PRIVACY PAPERS FOR POLICYMAKERS

2023

February 27, 2024

We are pleased to introduce FPF's 14th annual Privacy Papers for Policymakers. Each year we invite privacy scholars and authors to submit scholarship for consideration by a committee of reviewers and judges from the FPF Advisory Board. The selected papers are those judged to contain practical analyses of emerging issues that policymakers in Congress, in federal agencies, at the state level, and internationally will find useful.

This year's winning papers examine a variety of topical privacy issues:

- One paper analyzes the regulatory regime for international transfers of personal data in Latin America and proposes a regional solution.
- Another paper proposes that entities using algorithmic systems in traditional civil rights domains should have a duty to search for and implement less discriminatory algorithms.
- A third paper investigates the implications for public records governance of increased machine learning capability to generate sensitive inferences.
- The authors of another paper use interviews with experts on privacy-preserving methods and data sharing to highlight equity-focused work in statistical data privacy.
- Another paper critically examines how European countries are experimenting with facial recognition technologies in public spaces.
- A sixth paper describes how current laws fail to distinguish between inferences based on past conduct and algorithmic predictions about the future.
- The authors of another paper provide practical tips on how to conduct high-quality AI audits and discuss why audits are an essential component of responsible AI governance.
- Another paper examines the role of sensitive data in privacy legislation and contends that privacy law should focus on harm and risk rather than the nature of the data.
- The ninth paper tests current large language models' ability to infer sensitive personal information from input text and analyzes the privacy implications of such inferences.

For the eighth year in a row, we are proud to continue highlighting student work by honoring two papers: *The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government* and *Estimating Incidental Collection in Foreign Intelligence Surveillance: Large-Scale Multiparty Private Set Intersection with Union and Sum.*

We thank the scholars, advocates, and Advisory Board members who are engaged with us to explore the future of privacy.

Christopher Wolf
Founder and Board President,
FPF Board of Directors

Jules Polonetsky
CEO

Table of Contents

Awarded Papers

Honorable Mentions

Awarded Student Papers

Student Paper Honorable Mention

Out of respect for copyright law and for ease of reference, this compilation is a digest of the papers selected by the Future of Privacy Forum Advisory Board and does not contain full text. The selected papers in full text are available through the referenced links.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Future of Privacy Forum.

Towards a Latin American Model of Adequacy for the International Transfer of Personal Data

Luca Belli, Ana Brian Nougrères, Jonathan Mendoza Iserte, Pablo Palazzi and Nelson Remolina Angarita

Computers, Privacy and Data Protection Conference Latin America (CPDP LATAM) 2023

Available at: https://cpdp.lat/wp-content/uploads/2023/07/doc-discusion-cpdplatam23-2.3.pdf

Executive Summary

This article analyzes the regulatory regime for international transfers of personal data based on the legislation of several Latin American countries (namely Argentina, Brazil, Colombia, Mexico and Uruguay), its general regime and the different exceptions considered in the existing regulations. Finally, after explaining the divergences, different alternatives and ideas are proposed to create a specific regime to be used within Latin America for international transfers of personal data and recognition of adequacy. On the other hand, an analysis is carried out on the phenomenon of international data collection and solutions are proposed so that the rights of data owners are guaranteed when their information is collected from other countries without the collector being domiciled in the country of the data subject.

Authors

Luca Belli, Ph.D., is Professor of Internet Governance and Regulation at Fundação Getulio Vargas (FGV) Law School, where he heads the CyberBRICS project, and associated researcher at Centre de Droit Public Comparé of Paris 2 University. Luca is also a Member of the Board of the Alliance for Affordable Internet (A4AI), Director of CPDP LatAm and member of the CPDP Programme committee. Before joining FGV, Luca worked as an agent for the Council of Europe Internet Governance Unit and served as a Network Neutrality Expert for the Council of Europe.

Dr. Ana Brian Nougrères is a Professor of Law, Privacy and ICT at the University of Montevideo. She is based in Uruguay and is a practicing Attorney-at-law and Consultant on data protection. In July 2021, the Human Rights Council appointed her as the Special Rapporteur on the Right to Privacy and she took up the mandate on 1 August 2021. She presented her first UN report *Privacy and personal data protection in Ibero-America: a step towards globalization?* A/HRC/49/55 at the Human Rights Council in March 2022.

Jonathan Mendoza Iserte holds a PhD and a master's degree in Law from the Center for Graduate Studies in Law and a Law Degree from the National Autonomous University of Mexico. He has a master's degree in the General Data Protection Regulation of the European Union from the National Distance Education University of Spain and a certificate of aptitude for the specialization course "Cybersecurity Summer Bootcamp – Policy Makers", organized by the University of León, Spain, and the National Institute of Cybersecurity (INCIBE). Currently, Dr. Mendoza is Secretary of Protection of Personal Data of the National Institute of Transparency, Access to Information and Personal Data Protection (INAI).

Pablo A. Palazzi is a partner at Allende & Brea (Buenos Aires) with a focus on Internet law, data protection, Intellectual property, advertising law as well as IP litigation and technology transactions. He is an attorney licensed to practice in the State of New York and in Argentina and he is a cum laude graduate of the University of Fordham Law (NYC) where he earned a Master of Law degree in International Business and Trade Law. He is also President of the Argentine Chapter of Fordham Law School.

Nelson Remolina Angarita is a professor at the Law School of the University of the Andes (Bogotá, Colombia). He is the founder (2001) and director of GECTI (Group of Studies on the Internet, Electronic Commerce, Telecommunications, and Informatics) https://gecti.uniandes.edu.co/ and the Ciro Angarita Barón Observatory of Personal Data (2008) https://habeasdatacolombia.uniandes.edu.co/. He holds a Doctorate in Law (PhD) summa cum laude from Javeriana University (Bogotá, Colombia), a Master of Laws from the London School of Economics and Political Sciences, and is also a lawyer and specialist in commercial law from the University of the Andes. He served as the head of the Colombian data protection authority (from November 2018 to March 2022) and was the president of the Ibero-American Data Protection Network.

Less Discriminatory Algorithms

Emily Black, John Logan Koepke, Pauline Kim, Solon Barocas and Mingwei Hsu

Washington University in St. Louis Legal Studies Research Paper, Forthcoming

Available at: https://ssrn.com/abstract=4590481

Executive Summary

Entities that use algorithmic systems in traditional civil rights domains like housing, employment, and credit should have a duty to search for and implement less discriminatory algorithms (LDAs). Why? Work in computer science has established that, contrary to conventional wisdom, for a given prediction problem there are almost always multiple possible models with equivalent performance—a phenomenon termed model multiplicity. Critically for our purposes, different models of equivalent performance can produce different predictions for the same individual, and, in aggregate, exhibit different levels of impacts across demographic groups. As a result, when an algorithmic system displays a disparate impact, model multiplicity suggests that developers may be able to discover an alternative model that performs equally well, but has less discriminatory impact. But without dedicated exploration, it is unlikely developers will discover potential LDAs.

Model multiplicity has profound ramifications for the legal response to discriminatory algorithms. Under disparate impact doctrine, it makes little sense to say that a given algorithmic system used by an employer, creditor, or housing provider is either "justified" or "necessary" if an equally accurate model that exhibits less disparate effect is available and possible to discover with reasonable effort. As a result, the law should place a duty of a reasonable search for LDAs on entities that develop and deploy predictive models in covered civil rights domains. The law should recognize this duty in at least two specific ways. First, under disparate impact doctrine, a defendant's burden of justifying a model with discriminatory effects should be recognized to include showing that it made a reasonable search for LDAs before implementing the model. Second, new regulatory frameworks for the governance of algorithms should include a requirement that entities search for and implement LDAs as part of the model building process.

Authors

Emily Black is an Assistant Professor of Computer Science at Barnard College, Columbia University. Her work centers around creating methods to determine whether AI systems will cause various types of harm to the public, studying the equity impacts of AI systems in high-stakes settings, such as the government, and connecting her own and related research to the legal and policy worlds to help better regulate AI systems. She holds a PhD from Carnegie Mellon University, and is also an affiliate of Stanford's Reglab, where she completed her post-doc

Logan Koepke is a project director at Upturn, a nonprofit in Washington DC that advances equity and justice in the design, governance, and use of technology. His research and advocacy sits at the intersection of technology and civil rights. He helps lead Upturn's federal policy advocacy on the use of algorithmic systems in key civil rights contexts.

Pauline Kim is the Daniel Noyes Kirby Professor of Law at Washington University School of Law in St. Louis. An expert on the law of the workplace, Professor Kim has published numerous articles and book chapters on the impact of new technologies on workers' rights. Her research focuses on the risks of discrimination and unfairness posed by big data and artificial intelligence, and the legal and policy challenges posed by their adoption. Prior to law teaching, she clerked for the Honorable Cecil F. Poole on the United States Court of Appeals for the Ninth Circuit, then worked as a public interest lawyer in San Francisco, representing low-income workers facing discrimination, harassment, and illegal workplace conditions. She currently teaches employment discrimination law, the law of the workplace, and civil procedure at Washington University.

Solon Barocas is a Principal Researcher in the New York City lab of Microsoft Research, where he is a member of the Fairness, Accountability, Transparency, and Ethics in AI (FATE) research group. He's also an Adjunct Assistant Professor in the Department of Information Science at Cornell University, where he's part of the initiative on Artificial Intelligence, Policy, and Practice (AIPP). His research explores ethical and policy issues in artificial intelligence, particularly fairness in machine learning, methods for bringing accountability to automated decision-making, and the privacy implications of inference. He is co-author of the recently-published textbook on "Fairness and Machine Learning: Limitations and Opportunities" and he co-founded the ACM conference on Fairness, Accountability, and Transparency (FAccT).

Mingwei Hsu is a Senior Quantitative Analyst at Upturn. Upturn advances equity and justice in the design, governance, and use of technology. Mingwei's research focuses on a variety of areas, including credit and finance, labor and employment, and public benefits. Before Upturn, Mingwei was the V.P. of Enterprise Data and Analytics for Special Olympics. Prior to Special Olympics, Mingwei was a data scientist at the Consumer Financial Protection Bureau.

Future-Proofing Transparency: Re-Thinking Public Record Governance for the Age of Big Data

Beatriz Botero Arcila

Michigan State Law Review, Forthcoming

Available at SSRN: https://papers.ssrn.com/abstract=4535342

Executive Summary

Public records, public deeds, and even open data portals often include personal information that can now be easily accessed online. Yet, for all the recent attention given to informational privacy and data protection, scant literature exists on the governance of personal information that is available in public documents. This Article examines the critical issue of balancing privacy and transparency within public record governance in the age of Big Data.

With Big Data and powerful machine learning algorithms, personal information in public records can easily be used to infer sensitive data about people or aggregated to create a comprehensive personal profile of almost anyone. This information is public and open, however, for many good reasons: ensuring political accountability, facilitating democratic participation, enabling economic transactions, combating illegal activities such as money laundering and terrorism financing, and facilitating. Can the interest in record publicity coexist with the growing ease of deanonymizing and revealing sensitive information about individuals?

This Article addresses this question from a comparative perspective, focusing on US and EU access to information law. The Article shows that the publicity of records was, in the past and notwithstanding its presumptive public nature, protected because most people would not trouble themselves to go to public offices to review them, and it was practically impossible to aggregate them to draw extensive profiles about people. Drawing from this insight and contemporary debates on data governance, this Article challenges the binary classification of data as either published or not and proposes a risk-based framework that re-inserts that natural friction to public record governance by leveraging techno-legal methods in how information is published and accessed.

Author

Beatriz Botero Arcila is an Assistant Professor of Law at Sciences Po Law School and a Faculty Associate at the Berkman Klein Center for Internet and Society at Harvard University. She holds a doctorate from Harvard Law School where she defended a dissertation on the governance of smart cities and urban platforms and the data they collect.

Her research and work are on law and technology, with a focus on the governance of data and artificial intelligence. Recent work has focused on the governance of public data and public digital infrastructures, urban technologies, AI liability and the European Digital Strategy.

Beatriz is also co-founder of the Edgelands Institute, an organization focused on studying and creating local capacity on questions about digital surveillance and security in different cities around the world. Since its founding in 2021, the Edgelands Institute's work has spanned across cities in Colombia, Switzerland, Kenya and the US.

Previously, Beatriz worked in the environmental litigation team of Dejusticia, a think and do tank based in Bogotá, Colombia, and was head of legal of Tpaga, a digital payments company also based in Bogotá. Beatriz holds a law degree from the Universidad de los Andes, in Bogotá, Colombia and trained as a classical piano in the Universidad Javeriana in Bogotá.

Do No Harm Guide: Applying Equity Awareness in Data Privacy Methods

Claire McKay Bowen and Joshua Snoke

Available at: https://www.urban.org/research/publication/do-no-harm-guide-applying-equity-awareness-data-privacy-methods

Executive Summary

Researchers and organizations can increase privacy in datasets through methods such as aggregating, suppressing, or substituting random values. But these means of protecting individuals' information do not always equally affect the groups of people represented in the data. A published dataset might ensure the privacy of people who make up the majority of the dataset but fail to ensure the privacy of those in smaller groups. Or, after undergoing alterations, the data may be more useful for learning about some groups more than others. How entities protect data can have varying effects on marginalized and underrepresented groups of people.

To understand the current state of ideas, we completed a literature review of equity-focused work in statistical data privacy (SDP) and conducted interviews with nine experts on privacy-preserving methods and data sharing. These experts include researchers and practitioners from academia, government, and industry sectors with diverse technical backgrounds. We offer an illustrative example to highlight potential disparities that can result from applying SDP methods. We develop an equitable data privacy workflow that privacy practitioners and decisionmakers can utilize to explicitly make equity part of the standard data privacy process.

Authors

Claire McKay Bowen (she/her) is a senior fellow in the Center on Labor, Human Services, and Population and leads the Statistical Methods Group at the Urban Institute. Her research primarily focuses on developing technical and policy solutions to safely expand access to confidential data that advances evidence-based policy-making. She also has interest in improving science communication and integrating data equity into the data privacy process. In 2021, the Committee of Presidents of Statistical Societies identified her as an emerging leader in statistics for her technical contributions and leadership to statistics and the field of data privacy and confidentiality. Further, she is a member of the Census Scientific Advisory Committee and several other data governance and data privacy committees as well as an adjunct professor at Stonehill College.

Joshua Snoke is a statistician at the RAND Corporation. He researches statistical data privacy, fairness in machine learning, and workforce development. He utilizes statistical methodology to evaluate practical problems with the goal of developing better policy solutions. He received his Ph.D. in statistics with a graduate minor in social data analytics from Pennsylvania State University. He received his B.S. in mathematics and economics from Wheaton College.

Experiments with Facial Recognition Technologies in Public Spaces: In Search of an EU Governance Framework

Catherine Jasserand

An excerpt from: *Handbook on the Politics and Governance of Big Data and Artificial Intelligence (2023)*

Available at SSRN: https://papers.ssrn.com/abstract=4204452

Executive Summary

According to a survey conducted in 2020 by EDRi, at least 15 European countries have already used or experimented with facial recognition technologies (FRTs) in public spaces without much public debate. Yet, these highly intrusive technologies capture the distinctive facial characteristics of individuals to identify them. The systems operate at a distance without people's cooperation or awareness. Evidence from France and the United Kingdom shows that public authorities (mainly the police) have trialed and used the technologies in public spaces. Drawing insights from these experiments, the chapter assesses whether the applicable data protection frameworks are sufficient to regulate public authorities' experimentation with FRTs in public spaces. After identifying the regulatory gaps of the existing frameworks, the chapter provides some arguments and tools for a reflection on an experimental approach to test these technologies (such as Data Protection Impact Assessments, experimental legislation, and regulatory sandboxes based on the future AI Act).

Author

Catherine Jasserand has researched the intersection between law (fundamental rights) and biometrics for over 10 years. She did her PhD research in the area of privacy and biometrics at the University of Groningen (the Netherlands), focusing on the EU rules applicable to the law enforcement reprocessing of biometric data collected by private parties. Then, she pursued her postdoctoral research on the impact of facial recognition in public spaces on the EU rights to privacy and data protection as a Marie Curie recipient at the University of KU Leuven (Belgium) within CiTiP. She will join the University of Groningen to pursue research on AI, biometrics, and law.

The Prediction Society: Algorithms and the Problems of Forecasting the Future

Hideyuki Matsumi and Daniel J. Solove

Available at SSRN: https://papers.ssrn.com/abstract=4453869

Executive Summary

Today's predictions are produced by machine learning algorithms that analyze massive quantities of data, and increasingly, important decisions about people are being made based on these predictions.

Algorithmic predictions are a type of inference. Many laws struggle to account for inferences, and even when they do, the laws lump all inferences together. But predictions are different from other inferences and raise several unique problems. (1) Algorithmic predictions create a fossilization problem because they reinforce patterns in past data and can further solidify bias and inequality from the past. (2) Algorithmic predictions often raise an unfalsifiability problem. Predictions involve an assertion about future events. Until these events happen, predictions remain unverifiable, resulting in an inability for individuals to challenge them as false. (3) Algorithmic predictions can involve a preemptive intervention problem, where decisions or interventions render it impossible to determine whether the predictions would have come true. (4) Algorithmic predictions can lead to a self-fulfilling prophecy problem where they actively shape the future they aim to forecast.

More broadly, the rise of algorithmic predictions raises an overarching concern: Algorithmic predictions not only forecast the future but also have the power to create and control it. The increasing pervasiveness of decisions based on algorithmic predictions is leading to a prediction society where individuals' ability to author their own future is diminished while the organizations developing and using predictive systems are gaining greater power to shape the future.

Data protection/privacy law do not adequately address these problems. Many laws lack a temporal dimension and do not distinguish between predictions about the future and inferences about the past or present. We argue that the use of algorithmic predictions is a distinct issue warranting different treatment from other types of inference.

Authors

Hideyuki Matsumi, or Yuki, is a doctoral researcher at the Research Group on Law Science, Technology and Society (LSTS) of the Vrije Universiteit Brussel (VUB). He is also a member of the Health and Ageing Law Lab (HALL), and works on the EU H2020 project Hospital Smart development based on AI (HosmartAI). Member of the New York Bar.

His research focuses on issues at the intersection of law and data/information, specifically problems related to privacy, predictive analytics, and algorithms. Broadly, his research concerns how people are made vulnerable to automated decisions in a way that risks their autonomy and broader societal goals. At the narrowest level, he is now focusing on a pair of issues: (1) predictions or the temporal dimension of personal data, and (2) generated personal data and data brokers.

At the broader level, he has been addressing various problems related to information privacy, including international data transfers, information security, information accessibility, open-source software, digital abuse/assault/exploitation in the age of information, and genetic information.

He holds LLM in Intellectual Property Law from the George Washington University Law School, LLM with Law and Technology (IP) Certificate from University of California, Berkeley, School of Law, and LLM in International and European Law with specialization in Data Law from the VUB/IES (Great Distinction).

Prior to joining LSTS, he has worked in the academic sector as well as the business sector in Japan. He is a project associate professor at Keio University, and has worked before at the University of Tokyo and Toin University of Yokohama. He started his career as a programmer and information security consultant.

Daniel J. Solove is the Eugene L. and Barbara A. Bernard Professor of Intellectual Property and Technology Law at the George Washington University Law School.

One of the world's leading experts in privacy law, Solove is the author of 10+ books and 100+ articles. He has published books with Oxford, Harvard, and Yale University Presses, and articles in the Harvard, Stanford, and Columbia Law Reviews, among others. His works have been translated into many languages.

Solove founded two companies, one that provides privacy training to organizations and another that involves education, events, and certification to privacy professionals. He founded the Privacy Law Scholars Conference, the largest academic conference in privacy law. He served as co-reporter for the ALI's Principles of Law, Data Privacy.

A graduate of Yale Law School, Solove clerked for Judge Stanley Sporkin, U.S. District Court for the District of Columbia and Judge Pamela Ann Rymer, U.S. Court of Appeals for the 9th Circuit. He also was an associate at Arnold & Porter LLP and a senior policy advisor at Hogan Lovells LLP.

Solove has been interviewed and quoted in hundreds of media articles and broadcasts. He has more than 1 million LinkedIn followers. He has written a children's fiction book about privacy and has been depicted as a character in a play. He has been a consultant for many Fortune 500 companies and celebrities. He is the #1 most cited law professor born after 1970 and the #1 most cited law professor in the law and technology field.

AI Audits: Who, When, How... Or Even If?

Evan Selinger, Brenda Leong and Albert Fox Cahn

An excerpt from: Collaborative Intelligence: How Humans and AI are Transforming our World, (Forthcoming, MIT Press)

Available at: https://papers.ssrn.com/abstract=4568208

Executive Summary

Artificial intelligence (AI) tools are increasingly being integrated into decision-making processes in high-risk settings, including employment, credit, health care, housing, and law enforcement. Given the harms that poorly designed systems can lead to, including matters of life and death, there is a growing sense that crafting policies for using AI responsibly must necessarily include, at a minimum, assurances about the technical accuracy and reliability of the model design.

Because AI auditing is still in its early stages, many questions remain about how to best conduct them. While many people are optimistic that valid and effective best practice standards and procedures will emerge, some civil rights advocates are skeptical of both the concept and the practical use of AI audits. These critics are reasonably concerned about audit-washing—bad actors gaming loopholes and ambiguities in audit requirements to demonstrate compliance without actually providing meaningful reviews.

This chapter aims to explain why AI audits often are regarded as essential tools within an overall responsible governance system and how they are evolving toward accepted standards and best practices. We will focus most of our analysis on these explanations, including recommendations for conducting high-quality AI audits. Nevertheless, we will also articulate the core ideas of the skeptical civil rights position. This intellectually and politically sound view should be taken seriously by the AI community. To be well-informed about AI audits is to comprehend their positive prospects and be prepared to address their most serious challenges.

Authors

Evan Selinger is a professor of Philosophy at Rochester Institute of Technology. His research, writing, and teaching focus on ethical and legal issues related to technology, including privacy. He is a contributing writer at The Boston Globe and is currently writing a book for Cambridge University Press titled Move Slow and Upgrade: The Power of Incremental Innovation. The Observer selected his last book, Re-Engineering Humanity, as one of the best books of 2018. Evan has been a member of the Institute for Defense Analysis's Ethical, Legal, and Social (ELSI) working group for several years, advising DARPA on AI ethics. He has frequently collaborated with privacy organizations, including being a senior fellow at The Future of Privacy Forum and Scholar-in-Residence at the Surveillance Technology Oversight Project.

Brenda Leong is a partner at Luminos.Law (formerly BNH.AI), a boutique law firm uniquely founded by a partnership between lawyers and data scientists, dedicated entirely to developing policies and practices around AI governance, including applying Generative AI, building model risk management frameworks, performing model audits, and creating de-identification architecture and certification, along with designing and automating AI policies and procedures. Previously, Brenda was senior counsel and director of AI and ethics at the Future of Privacy Forum, where she oversaw the development and analysis of AI and ML resources. She is a recognized expert on the responsible use of biometrics and digital identity, with a focus on facial recognition, facial analysis, and emerging issues around voice-operated systems. Prior to her work at FPF, Brenda served in the US Air Force. She is a 2014 graduate of George Mason University School of Law.

Albert Fox Cahn is the Surveillance Technology Oversight Project's (S.T.O.P.'s) founder and executive director. He is also a Practitioner-in-Residence at N.Y.U Law School's Information Law Institute and a fellow at the Harvard Kennedy School's Carr Center For Human Rights Policy, Yale Law School's Information Society Project, Ashoka, and TED.

As a lawyer, technologist, and activist, Albert has become a leading voice on how to govern and build the technologies of the future. He started S.T.O.P. with the belief that local surveillance is an unprecedented threat to public safety, equity, and democracy.

Albert is a frequent commentator, with more than 100 articles in the New York Times, Boston Globe, Guardian, WIRED, Slate, NBC Think, Newsweek, and other publications. His TED Talk has been viewed hundreds of thousands of times. He frequently lectures at leading universities and speaks at leading technology governance forums. Albert previously served as an associate at Weil, Gotshal & Manges LLP, where he advised Fortune 50 companies on technology policy, antitrust law, and consumer privacy.

Albert also serves on the New York Immigration Coalition's Immigrant Leaders Council, IEEE Standards Association P3119 AI Procurement Working Group, and is an editorial board member for the Anthem Ethics of Personal Data Collection. He was also a founding member of the New York Immigrant Freedom Fund's Advisory Council. Albert received his J.D., cum laude, from Harvard Law School (where he was an editor of the Harvard Law & Policy Review), and his B.A. in Politics and Philosophy from Brandeis University.

Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data

Daniel J. Solove

Northwestern University Law Review, Vol. 118, 2024

Available at SSRN: https://papers.ssrn.com/abstract=4322198

Executive Summary

Heightened protection for sensitive data is trendy in privacy laws. Originating in EU data protection law, sensitive data singles out certain categories of personal data for extra protection. Commonly recognized special categories include racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health, sexual orientation and sex life, and biometric and genetic data.

Although heightened protection for sensitive data appropriately recognizes that not all situations involving personal data should be protected uniformly, the sensitive data approach is a dead end. The sensitive data categories are arbitrary and lack any coherent theory for identifying them. The borderlines of many categories are so blurry that they are useless. Moreover, it is easy to use nonsensitive data as a proxy for certain types of sensitive data.

With Big Data and powerful machine learning algorithms, most nonsensitive data give rise to inferences about sensitive data. In many privacy laws, data giving rise to inferences about sensitive data is also protected as sensitive data. Arguably, then, nearly all personal data can be sensitive, and the sensitive data categories can swallow up everything. As a result, most organizations are currently processing a vast amount of data in violation of the laws.

This Article argues that the problems with sensitive data make it unworkable and counterproductive as well as expose a deeper flaw at the root of many privacy laws. These laws make a fundamental conceptual mistake—they embrace the idea that the nature of personal data is a sufficiently useful focal point. But nothing meaningful for regulation can be determined solely by looking at the data itself. Data is what data does.

To be effective, privacy law must focus on harm and risk rather than on the nature of personal data. Privacy protections should be proportionate to the harm and risk involved with the data collection, use, and transfer.

Author

Daniel J. Solove is the Eugene L. and Barbara A. Bernard Professor of Intellectual Property and Technology Law at the George Washington University Law School.

One of the world's leading experts in privacy law, Solove is the author of 10+ books and 100+ articles. He has published books with Oxford, Harvard, and Yale University Presses, and articles in the Harvard, Stanford, and Columbia Law Reviews, among others. His works have been translated into many languages.

Solove founded two companies, one that provides privacy training to organizations and another that involves education, events, and certification to privacy professionals. He founded the Privacy Law Scholars Conference, the largest academic conference in privacy law. He served as co-reporter for the ALI's Principles of Law, Data Privacy.

A graduate of Yale Law School, Solove clerked for Judge Stanley Sporkin, U.S. District Court for the District of Columbia and Judge Pamela Ann Rymer, U.S. Court of Appeals for the 9th Circuit. He also was an associate at Arnold & Porter LLP and a senior policy advisor at Hogan Lovells LLP.

Solove has been interviewed and quoted in hundreds of media articles and broadcasts. He has more than 1 million LinkedIn followers. He has written a children's fiction book about privacy and has been depicted as a character in a play. He has been a consultant for many Fortune 500 companies and celebrities. He is the #1 most cited law professor born after 1970 and the #1 most cited law professor in the law and technology field.

Beyond Memorization: Violating Privacy Via Inference with Large Language Models

Robin Staab, Mark Vero, Mislav Balunovic and Martin Vechev

Available at: https://arxiv.org/abs/2310.07298

Executive Summary

Current privacy research on large language models (LLMs) primarily focuses on the issue of extracting memorized training data. At the same time, models' inference capabilities have increased drastically. This raises the key question of whether current LLMs could violate individuals' privacy by inferring personal attributes from text given at inference time. In this work, we present the first comprehensive study on the capabilities of pretrained LLMs to infer personal attributes from text. We construct a dataset consisting of real Reddit profiles, and show that current LLMs can infer a wide range of personal attributes (e.g., location, income, sex), achieving up to 85% top-1 and 95% top-3 accuracy at a fraction of the cost (100×) and time (240×) required by humans. As people increasingly interact with LLM-powered chatbots across all aspects of life, we also explore the emerging threat of privacy-invasive chatbots trying to extract personal information through seemingly benign questions. Finally, we show that common mitigations, i.e., text anonymization and model alignment, are currently ineffective at protecting user privacy against LLM inference. Our findings highlight that current LLMs can infer personal data at a previously unattainable scale. In the absence of working defenses, we advocate for a broader discussion around LLM privacy implications beyond memorization, striving for a wider privacy protection.

Authors

Robin Staab is PhD student at ETH Zurich SRI Lab, advised by Prof. Dr. Martin Vechev. Before this, he obtained both his Bachelor's and Master's degrees in Computer Science at ETH Zurich. He is the recipient of the Willi Studer Price for the Best Master's Degree in Computer Science in 2023. His research interests focus on the reliability and privacy of machine learning systems.

Mark Vero is a first-year PhD student at the Secure, Reliable, and Intelligent Systems Lab at ETH Zurich, under the advisorship of Prof. Dr. Martin Vechev. His research focuses on the security and privacy of modern machine learning systems, investigating federated learning, synthetic data, code generation, and large language models. He was part of the SRI team, winning the 2023 U.S. PETs Prize Federated Learning Red Teaming Challenge organized by NIST. Prior to his PhD studies, Mark completed his Bachelor's and Master's also at ETH Zurich, graduating as an Electrical Engineer.

Mislav Balunovic is a final year PhD student at ETH Zurich, supervised by Prof. Dr. Martin Vechev. Before that, he obtained his Master's degree where his Master thesis was awarded ETH Medal for Outstanding Thesis. The goal of his research is to develop robust, fair and private AI systems which would be fit for use cases which have high safety requirements. He also contributes to the open source projects which have the goal to increase AI Safety of large language models. Mislav was part of the winning team in the U.S. PETs Prize Red Teaming Challenge organized by NIST.

Martin Vechev is Full Professor of Computer Science at ETH Zurich where he leads the Secure, Reliable, and Intelligent Systems Lab. Born and raised in Sofia, Bulgaria he is also the Founder and Architect of INSAIT, the first world-class research center in computer science and artificial intelligence in Eastern Europe, created in partnership with ETH Zurich and EPFL. Prior to ETH, Martin was a Research Staff Member at the IBM T.J. Watson Research Center in New York, USA (2007–11). Before this he obtained his PhD from the University of Cambridge, England (2003–08).

Honorable Mention

The After Party: Cynical Resignation In Adtech's Pivot to Privacy

Lee McGuigan, Sarah Myers West, Ido Sivan-Sevilla and Patrick Parham

Big Data & Society, Vol. 10, 2023

Available at: https://journals.sagepub.com/doi/10.1177/20539517231203665

Executive Summary

Digital advertising and technology companies are resigned to a new privacy imperative. They are bracing for a world where third-party tracking will be restricted by design or by law. Digital resignation typically refers to how companies cultivate a sense of powerlessness about privacy among internet users. Our paper looks through this optic from the other end of the lens: How is the digital advertising industry coping with the increasing salience of privacy? Recent developments have forced companies to implement "privacy-preserving" designs—or at least promise some semblance of privacy. Yet, the industry remains dependent on flows of data and means of identification to enable still-desired targeting, measurement, and optimization. Our paper analyzes this contradiction by looking at systems that aim to replicate existing functionalities while protecting user "privacy." We call this a form of "cynical resignation" and characterize its key maneuvers as follows: (a) sanitizing surveillance; (b) party-hopping; and (c) sabotage. We argue that this "cynical resignation" to a privacy imperative represents a policy failure. In the absence of decisive interventions into the underlying business models of data capitalism, companies offer techno-solutionism and self-regulations that seem to conform to new laws and norms while reinforcing commitments to data-driven personalization. This may benefit the largest tech companies, since their privileged access to first-party data will make more companies reliant on them, and their computational power will be even more valuable in a world where modeling is used to compensate for the loss of third-party data and traditional methods of personal identification.

Authors

Dr. Lee McGuigan is an Assistant Professor in the Hussman School of Journalism and Media at the University of North Carolina at Chapel Hill. He is the author of Selling the American People: Advertising, Optimization, and the Origins of Adtech (MIT Press, 2023).

Dr. Sarah Myers West is the Managing Director of the AI Now Institute and recently served a term as a Senior Advisor on AI at the Federal Trade Commission. She holds a decade of policy and research experience in the political economy of the tech industry, and her forthcoming book Tracing Code (University of California Press) examines the origins of commercial surveillance.

Dr. Ido Sivan-Sevilla is an Assistant Professor in the College of Information Studies at the University of Maryland. His work bridges the computer science & public policy disciplines through the measurement and theorization of governance structures across a range of cybersecurity, privacy, and machine learning problems.

Patrick Parham is a Ph.D. candidate at the College of Information Studies, University of Maryland (UMD). He has been studying advertising and media technology, and proposals addressing the deprecation of third-party cookies. Patrick previously worked in the programmatic advertising industry.

Honorable Mention

Epsilon-Differential Privacy, and a Two-Step Test for Quantifying Reidentification Risk

Nathan Reitinger and Amol Deshpande

Jurimetrics Journal, Vol. 63, 2023

Available at: https://www.americanbar.org/content/dam/aba/publications/Jurimetrics/spring-2023/epsilon-differential-privacy-and-a-two-step-test-for-quantifying-reidentification-risk.pdf

Executive Summary

Sharing data in the twenty-first century is fraught with error. Most commonly, data is freely accessible, surreptitiously stolen, and easily capitalized in the pursuit of monetary maximization. But when data does find itself shrouded behind the veil of "personally identifiable information" (PII), it becomes nearly sacrosanct, impenetrable without consideration of ambiguous (yet penalty-rich) statutory law—inhibiting utility. Either choice, unnecessarily stifling innovation or indiscriminately pilfering privacy, leaves much to be desired.

This Article proposes a novel, two-step test for creating futureproof, bright-line rules around the sharing of legally protected data. The crux of the test centers on identifying a legal comparator between a particular data sanitization standard—differential privacy: a means of analyzing mechanisms that manipulate, and therefore sanitize, data—and statutory law. Step one identifies a proxy value for reidentification risk which may be easy calculated from an ε-differentially private mechanism: the guess difference. Step two finds a corollary in statutory law: the maximum reidentification risk a statute tolerates when permitting confidential data sharing. If step one is lower than or equal to step two, any output derived using the mechanism may be considered legally shareable; the mechanism itself may be deemed (statute, ε)-differentially private.

The two-step test provides clarity to data stewards hosting legally or possibly legally protected data, greasing the wheels in advancements in science and technology by providing an avenue for protected, compliant, and useful data sharing.

Authors

Nathan Reitinger is a late-stage PhD candidate at the University of Maryland in the Department of Computer Science where he works on problems lying at the intersection of law and computer science. He has written extensively on data sanitization, machine learning, cryptography, autonomous weapons systems, and intellectual property rights for 3D printing. He holds an M.S. from Columbia University and a J.D., magna cum laude, from Michigan State University. His work has appeared in both top computer science venues like IEEE Security and Privacy, the Privacy Enhancing Technologies Symposium, and USENIX, and law journals like the Stanford Technology Law Review, Jurimetrics, and American University Law Review.

Amol Deshpande is a Professor in the Department of Computer Science at the University of Maryland with a joint appointment in the University of Maryland Institute for Advanced Computer Studies (UMIACS). He received his Ph.D. from the University of California at Berkeley in 2004. His research interests include big data systems, collaborative data science platforms, machine learning lifecycle management, data privacy, graph analytics, and data lakes. He has authored or co-authored over 80 research publications, with an h-index of 51; he is a recipient of an NSF Career award, and has received best paper awards at the VLDB 2004, EWSN 2008, and VLDB 2009 conferences. He was also a Co-Founder and Chief Scientist at WireWheel, Inc., an enterprise provider of data privacy solutions (acquired by Osano).

Awarded Student Paper

The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government

Arushi Gupta, Victor Y. Wu, Helen Webley-Brown, Jennifer King and Daniel E. Ho

2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)

Available at: https://hai.stanford.edu/sites/default/files/2023-06/Gupta_et_al_Privacy_Bias.pdf

Executive Summary

An emerging concern in algorithmic fairness is the tension with privacy interests. Data minimization can restrict access to protected attributes, such as race and ethnicity, for bias assessment and mitigation. Less recognized is that for nearly 50 years, the federal government has been engaged in a large-scale experiment in data minimization, limiting (a) data sharing across federal agencies under the Privacy Act of 1974, and (b) data collection under the Paperwork Reduction Act. We document how this "privacy-bias tradeoff" has become an important battleground for fairness assessments in the U.S. government and provides rich lessons for resolving these tradeoffs. President Biden's 2021 racial justice Executive Order 13,985 mandated that federal agencies conduct equity impact assessments (e.g., for racial disparities) of federal programs.

We conduct a comprehensive assessment across high-volume claims agencies that affect many individuals, as well as all agencies filing "equity action plans," with three findings. First, there is broad agreement in principle that equity impact assessments are important, with few parties raising privacy challenges in theory and many agencies proposing substantial efforts. Second, in practice, major agencies do not collect and may be affirmatively prohibited under the Privacy Act from linking demographic information. This has led to pathological results: until 2022, for instance, the US Dept. of Agriculture imputed race by "visual observation" when race information was not collected. Data minimization has meant that even where agencies want to acquire demographic information in principle, the legal, data infrastructure, and bureaucratic hurdles are severe. Third, we derive policy implications to address these barriers.

Authors

Arushi Gupta is a master's student in Computer Science at Stanford University, where she also received her B.A. in Political Science. She is a research assistant at the Stanford RegLab, currently working on computer vision applications to help the EPA identify unpermitted facilities. She is interested in the intersection of data science with environmental and housing justice, and brings experience with policy research, quantitative analysis, and community-based advocacy to the space.

Victor Y. Wu is a student at Stanford Law School, where he serves as Managing Editor of the Stanford Law Review, Editor-in-Chief of the Stanford Environmental Law Journal, and Co-President of the Stanford Environmental Law Society. Before law school, Victor graduated as valedictorian from Dartmouth College, where he majored in Government, Environmental Studies, and Quantitative Social Science. In his free time, Victor enjoys triathlon, chess, rock climbing, and piano.

Helen Webley-Brown is a Ph.D. student in Political Science at MIT with specializations in American Politics and Quantitative Methods. Her current research agenda examines how artificial intelligence is impacting peoples' experiences and perceptions of government, especially within the criminal legal system. She completed her BA in Political Science at Washington University in St. Louis.

Dr. Jennifer King is the Privacy and Data Policy Fellow at the Stanford University Institute for Human-Centered Artificial Intelligence. An information scientist by training, Dr. King is a recognized expert and scholar in information privacy. Sitting at the intersection of human-computer interaction, law, and the social sciences, her research examines the public's understanding and expectations of online privacy as well as the policy implications of emerging technologies. Dr. King completed her doctorate in Information Management and Systems at the University of California, Berkeley School of Information.

Daniel E. Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science, Professor of Computer Science (by courtesy), Senior Fellow at Stanford's Institute for Human-Centered Artificial Intelligence, and Senior Fellow at the Stanford Institute for Economic Policy Research at Stanford University. He is a Faculty Fellow at the Center for Advanced Study in the Behavioral Sciences and is Director of the Regulation, Evaluation, and Governance Lab (RegLab). Ho serves on the National Artificial Intelligence Advisory Committee (NAIAC), advising the White House on artificial intelligence, as Senior Advisor on Responsible AI at the U.S. Department of Labor, and as a Public Member of the Administrative Conference of the United States (ACUS). He received his J.D. from Yale Law School and Ph.D. from Harvard University and clerked for Judge Stephen F. Williams on the U.S. Court of Appeals, District of Columbia Circuit.

Awarded Student Paper

Estimating Incidental Collection in Foreign Intelligence Surveillance: Large-Scale Multiparty Private Set Intersection with Union and Sum

Anunay Kulshrestha and Jonathan Mayer

31st USENIX Security Symposium (2022)

Available at: https://www.usenix.org/system/files/sec22-kulshrestha.pdf

Executive Summary

Section 702 of the Foreign Intelligence Surveillance Act authorizes U.S. intelligence agencies to intercept communications content without obtaining a warrant. While Section 702 requires targeting foreigners abroad for intelligence purposes, agencies "incidentally" collect communications to or from Americans and can search that data for purposes beyond intelligence gathering. For over a decade, members of Congress and civil society organizations have called on the U.S. Intelligence Community (IC) to estimate the scale of incidental collection. Senior intelligence officials have acknowledged the value of quantitative transparency for incidental collection, but the IC has not identified a satisfactory estimation method that respects individual privacy, protects intelligence sources and methods, and imposes minimal burden on IC resources.

In this work, we propose a novel approach to estimating incidental collection using secure multiparty computation (MPC). The IC possesses records about the parties to intercepted communications, and communications services possess country-level location for users. By combining these datasets with MPC, it is possible to generate an automated aggregate estimate of incidental collection that maintains confidentiality for intercepted communications and user locations.

We formalize our proposal as a new variant of private set intersection, which we term multiparty private set intersection with union and sum (MPSIU-Sum). We then design and evaluate an efficient MPSIU-Sum protocol, based on elliptic curve cryptography and partially homomorphic encryption. Our protocol performs well at the large scale necessary for estimating incidental collection in Section 702 surveillance.

Authors

Anunay Kulshrestha is an information security, applied cryptography, and technology policy researcher. He is a doctoral candidate in Computer Science at the Center for Information Technology Policy at Princeton University, where he is supported by a Wallace Memorial Fellowship in Engineering. Anunay develops novel privacy-enhancing techniques that improve accountability of digital systems integral to our social, political, and economic lives. His research attempts to combine robust accountability with strong cryptographic privacy. Anunay received his B.S. in Computer Science and Mathematics, and his M.A. in Public Policy from Stanford University.

Jonathan Mayer is an Assistant Professor at Princeton University, with appointments in the Department of Computer Science and the School of Public and International Affairs. He studies the intersection of technology and law, with emphasis on national security, criminal procedure, consumer privacy, network management, and online speech. Before joining the Princeton faculty, Jonathan served as the technology advisor to U.S. Senator Kamala Harris and as the Chief Technologist of the Federal Communications Commission Enforcement Bureau. Jonathan received his Ph.D. from the Stanford University Department of Computer Science and his J.D. from Stanford Law School.

Student Paper Honorable Mention

Ditching "DNA on Demand": A Harms-Centered Approach to Safeguarding Privacy Interests Against DNA Collection and Use by Law Enforcement

Emma Kenny-Pessia

Washington University Law Review, Vol. 101, 2023

Available at SSRN: https://wustllawreview.org/2023/12/27/ditching-dna-on-demand-a-harms-centered-approach-to-safeguarding-privacy-interests-against-dna-collection-and-use-by-law-enforcement

Executive Summary

Over the past decade, stories of questionable DNA collection and search methods by police have peppered the headlines. From this pattern of headline-making DNA collection practices and usages in law enforcement investigations emerges the grim reality of "DNA on Demand," where police have frictionless access to vast quantities of genetic information. With DNA collection and storage faster and cheaper than ever—and few legal rules circumscribing genetic data collection and use—law enforcement and private actors have embraced genetic data maximalism, assembling extensive, interconnected troves of intimate genetic information that may be searched and used indefinitely, even in ways completely attenuated from the initial DNA collection. While recent state laws targeting particular types of DNA database searches or collection methods evince a political appetite for genetic privacy protection, they are likely to be ineffective against the wide variety of DNA collection techniques that police have at their disposal, or against the assemblage of genetic data repositories that police may search and frictionlessly move between if access to one particular database type is shunted.

I suggest that only by appraising the severe privacy harms suffered across the wide variety of DNA collection methods and databases can we begin to view the many variations of genetic privacy violations as a single problem, which is a crucial first step in order to meaningfully safeguard privacy interests. By centering the discussion around the privacy harms wrought by DNA overcollection and overuse, the law can overcome its strictly procedural focus and move toward substantive and meaningful limits on law enforcement access to and use of genetic information.

Author

Emma Kenny-Pessia is a third-year J.D. Candidate at Washington University School of Law in St. Louis, Missouri. She has served on the editorial board of the Washington University Law Review, as a research assistant for Professor Neil M. Richards, as a student fellow with the Cordell Institute of Medicine and Policy, and as a student attorney in the WashU First Amendment Clinic. Emma's research interests include platform power, data brokers, and the relationship between privacy and free expression.

Before law school, Emma received a B.A. in Medicine, Literature, and Society from Barnard College, where she earned the Catherine Medalia Johannet Prize for her senior thesis which explored the illusion of "control" over individual privacy in the direct-to-consumer genetic testing sector.

Thank you to our 2023 Reviewers and Finalist Judges

Submissions received numeric rankings from a diverse team of academics, consumer advocates, and industry privacy professionals from the FPF Advisory Board, with each submission being evaluated for originality; overall quality of writing; and applicability to policymaking. For more information, visit fpf.org/privacy-papers-for-policy-makers.

Advisory Board Reviewers

Debra Berlyn
Project GOAL

Aaron Burstein
Kelley Drye & Warren

Hilary Cain
Alliance for Automotive Innovation

Sara Collins

Jacquie Cooke
23andMe

Michael Dolan
Best Buy

Eric Heath
Privacy Attorney

Joseph Jerome
University of Tampa

Mihir Kshirsagar
Princeton

Paul Lekas
Software & Information Industry Association

Aaron Massey
Future of Privacy Forum

Carlos Melvin
VP, Global Data Privacy Officer, Starbucks

Carolyn J. Pfeiffer
Janssen R&D

Teodora Pimpireva Tapping

Jason Sarfati
Chief Privacy Officer & VP Legal for Gravy Analytics, Inc.

Anne Toth
Privacyworks

Ron Whitworth
Truist

Farah Zaman
General Counsel, ElevenLabs

Cobun Zweifel-Keegan
International Association of Privacy Professionals

Finalist Judges

Bianca-Ioana Marcu
Policy Manager for Global Privacy, Future of Privacy Forum

Jules Polonetsky
CEO, Future of Privacy Forum

David Sallay
Director for Youth & Education Privacy, Future of Privacy Forum

Adonne Washington
Policy Counsel for Data, Mobility, Location, Future of Privacy Forum

PRIVACY PAPERS FOR POLICYMAKERS 2023

Future of Privacy Forum (FPF) is a nonprofit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies.

1350 EYE STREET NW | SUITE 350 | WASHINGTON, DC 20005 FPF.ORG | 202.220.8062