“The Future of Identity in Financial Services: Threats, Challenges, and Opportunities,”

The Task Force on Artificial Intelligence
House Financial Services Committee
United States Congress

Anne L. Washington, PhD
Steinhardt School
New York University

September 12 2019

Thank you for inviting me to speak today. My name is Anne Washington. I am an Assistant Professor of Data Policy in the Department of Applied Statisticsootnote{I am in the Steinhardt School of Culture, Education, and Human Development at NYU and the Department of Applied} at New York University. Before my career in Academia, I spent eight years in financial services with a data-driven company in San Francisco. I also spent a decade in the legislative branch working with many of the data structures and tools used to make this hearing possible. I would also like to acknowledge that I serve on the Academic Advisory Board of the Electronic Privacy Information Center, EPIC. I hold an undergraduate degree in computer science from Brown University, a graduate degree in Library and Information Science from Rutgers University, and a doctorate from The George Washington University School of Business.

My testimony, today, represents my own views as a public interest technologist. As a computer scientist and organizational scholar with expertise in open government data, I am part
of a growing movement of people\textsuperscript{2} using STEM\textsuperscript{3} skills in non-profits and the public sector. My academic specialty is understanding the organizational dynamics that shape the production and consumption of information, especially in organizations that have a public mission.

The courses I teach to graduate students at New York University are the “Management and Ethics of Data” and the “Ethics of Data Science”. In my testimony today, I will give you a crash course on data ethics, squeezing two semester-long courses into a five-minute briefing.

Artificial intelligence is not infallible. Even the most successful artificial intelligence systems used by online financial platforms require human input. For Americans to participate equally in our financial system, we need inclusive innovation that is aware of difference. Ignoring AI exceptions in financial services risks excluding many in our society because they are outliers from expectations. Organizations must begin to think about how they will handle future disputes over AI errors.

Artificial intelligence in the financial sector is an ethical, mathematical, and policy issue. To illustrate this, I will elaborate on three main points:

1. Artificial intelligence produces errors. When operating “at scale” even low error rates can impact millions. Errors in financial services will be consequential to specific individuals.

2. Because organizations are more likely to believe their technology systems over the experiences of individuals, individuals need procedures for recourse in the event of processing error.

3. Systems built to consider a broader range of populations must be more fault tolerant of cultural difference to be robust.

\textsuperscript{2} Such as Desmond Patton trained in computer science and social work at Columbia University. Dierdre Mulligan trained in Law and teaching in the Berkeley Information School. See Bruce Schneier’s Public Interest Technology list.

\textsuperscript{3} Science Technology Engineering and Math
Ethics

The study of ethics concerns itself with questions of appropriate behavior and actions. For centuries, the assumption behind ethics has been that we, as human beings, were driving our actions. Today, we are confronted with computer systems acting on behalf of humans. Ethical questions arise when actions violate the public trust.

Artificial intelligence is a technology that gives organizations an incredible power over individuals. M. Lynne Markus (2016) reminds us that the information on millions of people is in the hands of only a few and those organizations have a "corporate data responsibility." 

Data technology, such as artificial intelligence, drives all sectors of industry including financial services. Digital material from sensors, transactions, cell phones, networks, social media, and other digital traces feed into systems that generate artificial intelligence. Digital traces like these when reused in new contexts might trigger ethical concerns if not traceable and joined appropriately.

These pipelines into the "data supply chain" are mostly owned and operated by corporate bodies and not individuals. Christine Borgman (2015) argues that digital systems generate not just big data, but also small data, or even no data. These natural inconsistencies can create havoc when data technologists attempt to connect data from different sources. Many of these

---


technologies are perfectly legal and necessary for innovative growth, however ethical questions remain. Those in power who use this data must be reminded that data can make people vulnerable.

Artificial intelligence (AI) has grown our economy by driving economies of scale. Its efficiency provides gains in productivity and precision (Dhar, 2013; Halevey & Norvig, 2009). AI, however, can also obscure policies (Eubanks, 2018), and exacerbate bureaucracy (Peeters & Schuilenburg, 2018) amongst other concerns (Rossi, 2019; Wiggen, 2017). The tension between pragmatic efficiency and the moral tug of appropriate action plagues adoption of AI technologies by governments.

The artificial intelligence I discuss here, today, is data technology that enables oversight, automates decisions, or augments observations over large streams of data. Data technology includes data science, machine learning, predictive analytics, evidence-based policy, and computational tools based on the consumption and analysis of large quantities of information. Usually, these systems work with algorithms that sort, rank, search, and calculate in order to generate consistent outcomes. On the surface, these systems appear to be neutral, mechanical, and routine-driven, but when placed within human societies they can have substantial repercussions within our daily lives. The computer scientist Meredith Broussard says that these socially agnostic systems are not robust enough and labels them as "artificial unintelligence".9

The power of data technology is derived from the amount of data it uses. When data technology struggles to identify individuals in a database, the solution is often to combine more


databases into decision making. Amassing data in this way makes individuals entirely too visible (Rocher, 2019; Sweeney, 2013). Privacy and the "politics of real names,"10 danah boyd tells us, are real concerns. At the core of these concerns are questions not only of social categorization11 (Cherng, 2017) but also of technical abstraction (Walsh, 1992).

Ethics programs, like the one at NYU, want to help build better data systems. By baking privacy, security, and usability into the design of our AI systems, we can build a more responsible and ethical data environment like the solutions proposed by (Shilton, 2013) and (Cranor & Garfinkel, 2005)12. Others, such as the scholars at the Ostrom Center for Data Commons, are using the work of Nobel-prize winning economist, Elinor Ostrom, to better understand the ethics of knowledge commons (Raymond, 2018).

Data Ethics In Real Life

Every student of data ethics understands that large populations, coincidence, and cases of mistaken identity can confound the “trustworthiness” of AI systems.

Large Data Sets

Current AI-based systems, including financial systems incorporating AI, are not ready to

disambiguate enormous sets of people. The USA currently has close to 330 million people. Some technology platforms have more users than the populations of countries. These large Internet platforms have specific authentication methods to identify a user who logs in that includes active involvement of the individual logging in. Financial service data, which travels between multiple institutions, is harder to track down and does not include the end-user actively as part of authentication and identification. This leads to loosely coupled systems that introduce noise into financial profiles. Evelyn Ruppert (2009, 2011) calls the digital traces that represent us a “data double”. Our data double is similar to us but not exactly. Daniel Solove (2004) calls these data traces an “unauthorized biography” that contains some true things but lots of noise and innudendo that is not true. The scale of users and their data can introduce error into the decisions made by financial sector AI.

The Birthday Problem

Coincidences\textsuperscript{13} are not a surprise to any student of statistics. Basic math theory tells us that what we expect is rare may be more likely than we think. The inquiry known as the Birthday Problem\textsuperscript{14} asks: how many people are needed in a room for a good chance that two people are born on the same day of the year? Surprisingly the number is just 23. There is a 50\% chance that in a room of 23 people, given true randomness, two people have the same birthday. In a room of 75 people, the chances are over 99\% and a 1/3 chance that three people do. When two items resolve to the same set of information, computer scientists building a hash algorithm know this as a “hash collision”.

The Birthday Problem has real world implications when we use this information to

\textsuperscript{13} Stewart, I. (1998). What a Coincidence! - MATHEMATICAL RECREATIONS. Scientific American, 2. doi: DOI: 10.1038/scientificamerican0698-95

disambiguate identities. A classic example is the problem of watch lists that permit entry or deny services. Jeff Jonas, a pioneer in entity recognition, has explored this tension between privacy and recognition in a famous paper about the terrorist watch list. The authors conclude that actionable information is more important than aggregate lists that violate civil liberties (Harper & Jonas, 2006). The legal scholar Margaret Hu (2015) goes into extensive detail about these lists and their impact on people's lives.

These problems are not new (Solove, 2001; 2004) nor unknown to computer scientists and statisticians (Becker, 2006). What is new, however, is that these materials are moving from identification into action in ways that can aggregate a single mistake into an ongoing situation. The data supply chain moves not in one direction but in circles exacerbating mistakes.

Examples of mistaken identity

People have a difficult time fighting these lists once their names are on them.

• Jennifer Norris\(^{15}\) of Boston was in danger of losing her job because of the inability to resolve a dispute about her identity. Her work required a driver’s license and only after consulting her Congressman, Capuano of Massachusetts' 7th district and a local news agency was the problem resolved.

• Kathleen Casey\(^ {16}\), a pharmacy technician, lost her apartment in 2011 when a system confused her with someone else. It is important to note that some industries\(^{17}\) such as retail pharmacy stores used informal lists to exclude any job candidate accused of theft.

---


• A teacher in Maryland\textsuperscript{18} could not pursue her chosen livelihood because bad data continually haunted her in a job that requires continuous recertification.

The astounding case of Lisa S. Davis\textsuperscript{19}, the novelist, who wrote about her experience of encountering her data double in official documents for 18 years and finally meeting her. For years, their addresses were confused and they would get mail for each other. They had the same day of birth, the same year of birth, and not only the same middle initial, but the same middle name. Most systems have a hard time if not impossible time disambiguating them. They assume it is one person who perhaps has just moved to a new address. This data double story has more resonance in this case because the two women are different colors and live in neighborhoods with different policing behaviors. They are both in New York State so their information has a higher chance to be co-mingled in databases.

**Resolving disputes**

Davis (2017) relates her story of having information that would show that her experience and paper traces verified who she was. Her lived experience was no match to the certainty of a computer. She was assumed to be a liar and told to plead guilty to pay and clear the traffic violations.

Organizations tend to trust their computer systems over the customers’ experience. Individuals with a wrong match, who are outliers, who clearly can identify a flaw in the system, are perceived as liars. Humans take the blame after a systems provides an answer. People with


\textsuperscript{19} Davis, Lisa Selin (2017, Apr 3) For 18 years, I thought she was stealing my identity. Until I found her . The UK Guardian https://www.theguardian.com/us-news/2017/apr/03/identity-theft-racial-justice
lived experience that contradicts the artificial intelligence face significant challenges. It is like watching a toy robot go towards the corner and march in place endlessly.

Technologists building these systems want to learn this feedback. Businesses do not have a financial incentive to incrementally fix small errors. Any policy or best practice would give technologists inside organizations the leverage they need to spend their time fixing the errors. This feedback once incorporated could help prevent similar mistakes from being repeated later. This agile approach with feedback would help to incrementally improve the technology.

Individuals should have recourse in these situations. It is mathematically certain that collisions will occur. Without any form of redress, innovation will stall and many people will be locked out of financial systems.

It is important to note that these stories all focus on individuals but one-person Internet shops\textsuperscript{20} that rely on technology infrastructure are even more vulnerable. Owner-operator and new entrepreneurs, and small business who are establishing their validity in markets have high risks if locked out of financial capital.

\textbf{POLICY}

My remarks on policy alternatives will be the most brief. Legal scholarship in this area is extensive especially in data\textsuperscript{21} used in policing and court data\textsuperscript{22} system. The scholarship of

\begin{footnotesize}
\begin{itemize}
\end{itemize}
\end{footnotesize}
Danielle K. Citron and Frank A. Pasquale have covered many plausible solutions.

I suspect that there will be a debate over the feasibility of establishing recourse. As usual, some will suggest that new regulations should be put in place. It would be logical to extend the Fair Credit Reporting Act (FCRA) into the 21st century and acknowledge the role of data sources. Others will suggest that self-monitoring would be sufficient. It makes sense to allow innovation to develop without unnecessary constraints since the future of these technologies is hard to foresee. Associations and industry cooperatives could continue to establish best practices across the field.

In my opinion, neither of these traditional responses gets to the heart of the issue which is that data-driven organizations need to establish internal data policy that matches their values with the business model. Data-driven organizations might run on a variety of business models (Shapiro & Varian, 1998) so it is a management decision what policy best matches those goals. I see these concerns as extensions of early conversations about organizational memory (Ackerman, 2000; Anand, 1998) that asked what information should organizations keep as technology became ubiquitous.
One critical role for public policy is data standards. Governments could establish data standards that would relieve the burdens of anyone trying to investigate issues across firms. Standardized data structures could also a mechanism to trigger retrospective tracking for regulators, e-discovery, or internal business intelligence. The reuse and exchange of digital material is complicated by many social and organizational challenges (Borgman, 2000; Bowker, 1996; Edwards, 2011; Fedorowicz, 2010; Markus, 2006). A solid internal information policy (McClure 1989; Robinson, Yu, Zeller, & Felton, 2008) is critical for any data-driven organization. For example, in the public sector the 2014 DATA Act produced a stable data infrastructure across all agencies that made later analysis, correction, and innovation possible. Digital government scholars such as Sharon Dawes (1996, 2010), Theresa A. Pardo (2012), Marijn Janssen (2016), Lemuria Carter (2018), Paul Jaeger & John Bertot (2010) have written extensively about the importance of data structures in government transparency.

Governments often neglect that their greatest power in public policy is mandating data infrastructure. Identity standards for financial services would greatly serve to expedite the adoption of artificial intelligence that benefits wide audiences.

**Summary**

Artificial intelligence, often implemented to save labor costs, will still require human labor to handle anticipated exceptions. A dispute resolution process solves two problems: procedural justice and technology improvement. First, it establishes a procedure to preserve the sanctity of human experience in situations where organizations may be more likely to trust the AI over a customer. Second, it provides the necessary feedback for incremental improvement of the technology.
Artificial intelligence will have its exceptions and people need procedures to assert the authority of their lived experience over the authority of the numbers.

BIBLIOGRAPHY


Chupapados Data https://chupapados.codingrights.org/en/they-are-stalking-you-to-calculate-your-score/


Davis, Lisa Selin (2017, Apr 3) For 18 years, I thought she was stealing my identity. Until I found her . The UK Guardian https://www.theguardian.com/us-news/2017/apr/03/identity-theft-racial-justice


Acknowledgements
Congressional testimony has one voice but many hands.

This would not have happened without the extraordinary voluntary support from my former doctoral research assistants: David C. Morar and Rachel S. Kuo. Someone should hire them soon because they are both outstanding talent. Molly Nystrom and Kiran Samuels identified the cases used in the testimony. It was delightful to reconnect to old friends from WFNIA / BGI / BlackRock through Janice Deringer. I am grateful for my current colleagues in the Applied Statistics program and the A3SR 'Data for Social Impact' students especially the Spring 2019 Ethics of Data Science class.  I appreciated the support from my fellow fellows at D&S, as well as the #unsettle event, The Hellenic Express, Team Rosario, and S.C.A.N. The Critical Race + Digital Scholars cooperative helped me identify additional citations for the bibliography.

All mistakes are my own, but I would like to acknowledge the expertise of the following people who contributed to this effort: David Morar, Rachel Kuo, Kiran Samuels, The Qlab, Jessica Spencer, Molly Nystrom, Chloe Marten, Shannon Kay, Alexandra Mateescu, David G. Robinson, Rashida Richardson, Meredith Broussard, Danielle Citron, Janice Deringer, Jenn Holmes, James R. Brandon, KO.T.N. , Mr. & Mrs. Merlin, Thing1 and Thing2.

anne.washington@nyu.edu
FSCTestimony Washington SUBMIT2.docx