Statement by Seny Kamara

Associate Professor
Department of Computer Science
Brown University

Chief Scientist
Aroki Systems

before the

Task Force on Financial Technology

of the

Committee on Financial Services
U.S. House of Representatives
November 21, 2019

Chairman Lynch, Ranking Member Emmer and distinguished members of the Task force on Financial Technology. I appreciate the opportunity to testify at today's hearing on the role of big data in financial services. Today, I will speak about how data is transforming the financial industry and how this transformation holds great promise but—unless it is carefully guided—also has the potential to erode consumer privacy and increase discrimination.

Experience. I am an Associate Professor of Computer Science at Brown University, where I conduct research in cryptography: the mathematical science that underlies data privacy and security. I am an affiliate of the Brown Data Science Initiative and of the Brown Center for Human Rights and Humanitarian Studies. Prior to Brown, I was a research scientist at Microsoft Research working in the Cryptography Research group. Over the last 20 years, I have developed a number of encryption algorithms and cryptographic protocols for data protection and privacy.

Overview. The financial industry is being transformed by technology. Examples include mobile devices, the Internet of Things (IoT), blockchains, smart contracts and machine learning. Both traditional institutions and technology startups are leveraging these technologies to provide new financial services to consumers. These developments can provide great benefits to consumers. Benefits that include expanding credit to the underbanked, offering better insurance rates to homeowners and improving fraud detection. While I want to recognize

the importance of these outcomes, it is critical that consumers and lawmakers understand the tradeoffs that these new financial technologies require. As is often the case when technology "disrupts" an industry, Fintech has the potential to both improve and harm the lives of people. These harms include the erosion of privacy and new forms of "algorithmic" discriminatory and predatory practices.

How Big Data is used in Financial Services

Big data usually refers to massive datasets and the systems and algorithms used to store, manage and analyze them. The data we produce—and is produced about us—is expected to grow from 29 zettabytes in 2018 to 175 zettabytes in 2025 [3]. While most conversations about big data focus on its size, an important dimension of data that is often overlooked is its *type*.

Types of data. Data comes from a variety of sources and is produced for a variety of reasons. For the purposes of this discussion, I will characterize data into three categories. The first is *authored data* which is produced by people. This includes emails, messages, tweets, comments and documents. The second category is *observational data* which is data that is produced about people by third parties; be it other people or algorithms. This includes, for example, medical records produced by physicians, consumer credit data produced by lenders, location data produced by mobile devices and automotive data produced by IoT devices in cars. Finally, there is *meta data* which is data that describes other data; for example, the date and time a piece of data was created and who the author was. It is important to highlight that *all* these data types are sensitive, not only authored data. In fact, it is by now well understood in the privacy research community that "all data is personally identifiable information". This is the case because even innocuous looking data about an individual can be correlated with her identity [9].

Data sources. The Financial industry is using new data sources, including authored, observational and meta data. These new sources, often called *alternative* data, range from utility bills to location data and text messages. For example, credit reporting agencies like Experian, TransUnion and Equifax are using data about "every day bills" to create new credit scores. Insurance companies and startups are using IoT data from homes and cars to better predict risk. In the past, some insurance startups tried to use Facebook posts and psychometric tests to asses people's risk profile [6]. Some mobile lending apps track location to determine how much time their users spend at work [2]. New micro-lending apps are using location data, social media content, contact lists and the behavior of Facebook friends to estimate people's credit-worthiness. An app made in California that operates in Kenya, even accesses call history under the belief that people who regularly call their mothers are more likely to repay their loans [5].

Collection and storage. Some of these Fintech apps have privacy policies that are vague and unfavorable to consumers. The data they collect is intrusive and sensitive and their terms of service effectively grant app developers ownership of customer data. Furthermore, data collection often occurs in the background even when the app is not in use and the collected data is stored and analyzed on company servers even after the app has been deleted [5].

Data processing. In addition to leveraging new sources of data, the financial technology industry is also processing data in new ways using machine learning models to make automated decisions quickly and at scale. While classical algorithms are designed by domain experts and expressed by a series of rules and explicit choices, machine learning models are produced by *algorithms* that learn from data. The models produced in this manner can be very effective in certain contexts but suffer from important limitations. The first is a lack of transparency: we often do not know and, therefore, cannot explain why a machine learning model makes a particular decision. This is a serious concern in the context of credit since the Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) require creditors to explain the reason an application was denied. The second important limitation of machine learning models is bias in decision making. While this kind of algorithmic discrimination has been well-publicized in the last few years, it is important to note that we are only in the very early stages of rigorously understanding this behavior of algorithms. In fact, in this space, there are currently more questions than answers so it is important to tread carefully. One thing we do know is that simply ignoring protected attributes like race and gender in machine learning is not enough to guarantee unbiased decisions [1] but some Fintech companies claim exactly this [11]. This is a serious concern in the context of the Equal Credit Opportunity Act and the Fair Housing Act, both of which prohibit discriminatory lending practices.

Privacy Laws and Financial Data

The privacy of financial records is governed by the Gramm-Leach-Bliley Act (GLBA), the Bank Secrecy Act, the Right to Financial Privacy Act and the FCRA. It is important to note, however, that these laws apply to financial records but that is not the entirety of the data a financial institution collects. Here, strong privacy laws like the California Consumer Protection Act (CCPA) fill an important gap left open by existing laws. Also, as new financial services and companies emerge, it may be difficult to ascertain whether they qualify as financial institutions as defined by pre-existing law. Filling this gap is critical and the Financial Information Data Modernization Act (FIDMA) clarifies uncertainties in the GLBA while providing strong protections for consumers with an eye towards to advances not only in financial technology but privacy technologies as well.

Innovations in Privacy Technologies

Fintech apps can make use of multiple sources of consumer data, ranging from financial records provided by a bank to location data provided by a mobile device. Traditionally, financial apps have shared data through a practice called *screen scraping*, where an app asks a user for their credentials (i.e., login and password) so that it can log into the user's accounts on its behalf and retrieve the information it needs. It is widely accepted that this practice is substandard from a privacy and security perspective since users have to completely trust the app to store, protect and not abuse its credentials.

APIs. A better approach, which is now being developed by the financial industry is to use APIs. Roughly speaking, APIs are standardized interfaces between apps that allow for easier inter-operability and improved security. With an API-based design, apps can access user data only through a user-approved token that determines which pieces of data can be accessed and for how long. APIs are a considerable improvement over screen scraping but they are far from enough to guarantee consumer privacy. With an API-based design, apps can still access, lose, exploit and abuse raw user data. And as long as consumers have to trust "data hungry" apps that scour their sensitive data under vague privacy policies, they will never have real privacy.

New privacy technologies. But what if consumers did not have to give up their data in order to benefit from financial and technological innovations? What if financial apps and services never had to see raw data? This might sound impossible but, in fact, it is! Over the last 30 years, cryptography researchers in academia and in industry labs have developed a wide array of cryptographic techniques to process encrypted data. This gives us the ability to run algorithms (including machine learning algorithms) over encrypted data, to search through encrypted files and to query encrypted databases—all without ever decrypting the data. This set of privacy technologies, which include secure multi-party computation, private set intersection, homomorphic encryption and encrypted search algorithms, can enable truly private data processing [4]. I want to stress here that these technologies are not science fiction; they are ready for use today. In fact, in 2017 the Boston Women's Workforce Council and Boston University deployed secure multi-party computation to privately analyze the wage gap in the greater Boston area [8]. This year, Google announced its deployment of private set intersection to privately process data with external partners [7]. And encrypted search algorithms are starting to be deployed by major database companies [10]. By leveraging these advances in cryptography, financial technologies could deliver on their promise to improve the financial health of their customers without them having to sacrifice their privacy.

The financial industry is being transformed by technology. And in the wake of this transformation it is easy to get carried away on a wave of technological optimism. As a computer

scientist, I believe in the power of technology but I am also acutely aware of its potential harms. As a cryptographer, I worry deeply about the erosion of privacy that these financial apps and services can cause. We are all aware of the constant occurrence of data breaches; of the weaponization of private data to micro-target people and affect their behaviors. Do we want another Equifax? Do we want another Cambridge Analytica? "Moving fast and breaking things" is not sound engineering practice and it is not sound policy. It is imperative that we proceed carefully and that we oversee this transformation with strong privacy laws and strong privacy technologies.

Thank you. I look forward to answering your questions.

References

[1] Solon Barocas and Andrew Selbst. Big data's disparate impact. *Calif. L. Rev.*, 104:671, 2016.

[2] Branch. How does branch determine my advance limit? https://support.branchapp.com/hc/en-us/articles/360029167251-Why-aren-t-my-hours-tracking-.

[3] John Rydning David Reinsel, John Gantz. The digitization of the world from edge to core. *An IDC White Paper-#US44413318*, 2018.

[4] Nigel Smart (ed), David Archer, Dan Bogdanov, Alexandra Boldyreva, Seny Kamara, Florian Kerschbaum, Yehuda Lindell, Steve Lu, Jesper Buus Nielsen, Rafail Ostrovsky, Jakob Pagter, Ahmad-Reza Sadeghi, and Adrian Waller. Future directions in computing on encrypted data, 2015.

[5] Privacy International. Fintech: Privacy and identity in the new data-intensive financial sector, 2017.

[6] Privacy International. Social media intelligence and profiling in the insurance industry: It' s not only the price you pay that will be affected. 2017.

[7] Mihaela Ion, Ben Kreuter, Ahmet Erhan Nergiz, Sarvar Patel, Mariana Raykova, Shobhit Saxena, Karn Seth, David Shanahan, and Moti Yung. On deploying secure computing commercially: Private intersection-sum protocols and their business applications. *IACR Cryptology ePrint Archive*, 2019:723, 2019.

[8] Andrei Lapets, Frederick Jansen, Kinan Dak Albab, Rawane Issa, Lucy Qin, Mayank Varia, and Azer Bestavros. Accessible privacy-preserving web-based data analysis for assessing and addressing economic inequalities. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, page 48. ACM, 2018.

[9] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18-21 May 2008, Oakland, California, USA*, pages 111–125. IEEE Computer Society, 2008.

[10] Lily Hay Newman. A plan to stop breaches with dead simple database encryption. *Wired*, 2019.

[11] Tala. Data ethics & consumer protection. https://tala.co/data-ethics/.