Privacy Protection vs. Prevention of Biases in Telecom Industry Data Virgin Territories Exploration

Hongyi Huang[1], Josh Matthews[2], Don Riley[3]
[1]Apkudo, Baltimore, MD bridget.huang@apkudo.com [2]Apkudo, Baltimore, MD josh.matthews@apkudo.com [3]Apkudo, Baltimore, MD don@apkudo.com

ABSTRACT

Several years ago, when customized services were on the forefront of everyone's mind, there was little indication that such customized services would lead to fears of systemic data privacy invasion. With the introduction of regulations like the EU General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA), the public now pays greater attention to their data privacy, which increases the tendency of users to pull their individual data out of the collective pool[1]. However, considering that the real-world data itself is not perfect, will this new trend from clients jeopardize the business statistical analysis quality by introducing more biases and creating more statistical limitations? The telecom industry can give a practical example. Exchange Risk Index ("ERI") as the new metric; the ERI adopts risk control methodology to optimize the quality monitoring of the mobile devices exchange process. i.e. it gives the quantified and normalized ability to estimate the risk performance of triage issues of certain mobile models. In addition, it enables the prioritization of the most critical issues and predicts future risk performance. Mobile devices exchanges consist of individual decision making. Of these "apple to orange" instances, each data point is critical. How much real-world data can businesses afford to lose through data privacy protections before data-supported decision making is impacted? How can businesses adjust their data operations and utilization strategies to adapt to the new norm? This is the new challenge we face today.

INTRODUCTION

Like many other industries, the telecommunications (hereafter referred to as "telecom") industry has been trying to improve the user experiences and customer services quality by meeting up with the various client requirements and offering customized services. Along with the fourth industrial revolution, to back up the ideology of optimizing the services quality and supply chain portfolio, the data becomes the new oil and the most important strategic resources. The Top Four US telecom companies have operated in the market for decades: AT&T (American Telephone and Telegraph) founded in 1983, Verizon (Verizon Communications Inc.) founded in 1983, T-Mobile US founded in 1990, and Sprint (Sprint Corporation) founded in 1986[2]. If utilized properly, the data cumulated in their legacy systems can support their success in the fourth industrial revolution on two levels: On the macro level, it can clearly map the networks among OEMs (original equipment manufacturer), carriers, vendors, distributors, wholesalers, retailers, and end customers in the supply chain; on the micro level, it can plot the comprehensive digital twin of each end customer based on personal identity info, finance info, geographic info, and call history.

Currently, usually company itself would not have only one centralized data storage warehouse. Due to the entanglement of the legacy systems and the complexity of the organization structure, it is very common to see the silos of data and even inconsistency of the data from different sources. When companies conducting data analyses or machine learning research, their data resources would not be limited to its own data. Purchasing data from third parties, scraping data from social media websites with significant customers flow, surveying targeted customers group are popular ways to enrich the raw datasets.

With the potentials of the data power shown by the artificial intelligence and machine learning, shutting power in the cage is becoming an emergency. The direction of this action can come from either internal such as self-regulation, or external such as governmental policies and regulations.

DATA PRIVACY

On July 12th, 2019, the historically highest $5 billion penalty to Facebook Inc. was approved by FTC (Federal Trade Commission) to resolve the Cambridge Analytica data scandal[3]. Considering the closeness of smart devices and the internet services to our modern life, it the telecom industry should be one of the first responders to news like this. It gives the telecom industry a warning that some actions should be taken before the fire spreading. Self-regulation is the necessary first solution, since it is faster, more detail-oriented, and easier to be implemented by the company itself.

Zuckerberg also suggests a series of key concepts for “Privacy-Focused Vision for Social Networking” with private interactions, encryption, reducing permanence, safety, interoperability, and secure data storage. Just like the Maslow’s hierarchy of needs demonstrates the human innate motivations priority, the data privacy needs also have its priority. If we generalize Zuckerberg’s key concepts and migrate it to the telecom industry, the data privacy pyramid would meet up the criteria: Static data security (data storage, reducing permanence), dynamic data security (interoperability), encryption, disaggregation from personal identifiable information (PII) data[4].

In the US market the data privacy regulations literally guarantee enabling customers to opt out with request and have the right to download their own data. However, a similar problem that the telecom industry shares with the social media is that opting out could be overly difficult considering how long and deep the telecom services have embedded in the customer’s daily usage.

Currently just like many other businesses, the telecom industry stays with the business model more revenue driven than cost containment driven. Even though the uncovered value of the reverse logistics is noticed by the industry itself, the focus and the efforts have not been totally turned toward. Revenue driven business model can benefit the telecom companies in the short term, but meanwhile it leaves data security concerns behind. For example, after factory data reset (FDR) [5,6] most customers would assume their data has been wiped completely before this device is handed to the carrier or a second hand phone buyer. However, this is not the truth. FDR only encrypts the data rather than removes it for good. The reason of this design is because carriers wish to save the time their store representative taking care of each phone return or exchange, which apparently does not drive their revenue. Wiping data completely takes them 15-30 minutes, while encrypting data only takes 5 minutes or less. After getting back to the retail store, this phone needs to wait until reaching return center to get its data completely removed. This step is supposed to be completed before product triaging, repairing, dismantling, or disposing steps.

To lower the risk of customer data breach, devices flow into the reverse supply chain need to be wiped in an earlier stage with data protection plan designed by domain experts. And it is recommended that this process can be monitored and audited by third-party data protection institute. On the other hand, if the device gets traded with no proper wiping in the second hand market, this self-insured data protection would cause higher risk by endangering seller's privacy. An experiment done by Security firm Avast [7] to prove the self-insured approach is no a smart data security solution: This firm purchased twenty previously-owned Android phones on eBay. With simple extraction methods available publicly, they recovered 40,000 old emails, texts, photos, and even the personal identities of the previous phone owners. The industry is trying to improve the data privacy protection by providing solutions like homeomorphic encryption, differential privacy, privacy-preserving machine learning, and zero-knowledge proofs.

The raw end-customers' data matters directly to end-customers' privacy. When discussing on data privacy protection, usually the current governmental regulations and policies are more focusing on the scope of the individual end customer's benefits. The regulations and policies are trying to open up the black boxes where the customer data flows in. However, these black boxes also matter to the confidentiality of business. If customers require for downloading their data collected by the business and getting explanation how the business utilizes their data, the business might not able to give a comprehensive introduction without leaking their business model kernels, including machine learning algorithms, and in confidentiality clause (also referred to as a nondisclosure agreement or NDA) the confidential information cannot be disclosed without proper authorizations. We can imagine there would be some potential conflicts between end customer data privacy protection and business confidentiality protection.

Another possible data privacy conflict point would merge from company employees. With the increase use of smart devices, more work-related content or data would be unintentionally pulled to unprotected personal devices. Inevitably, "bringing your own devices" (BYOD) to work or mixed use of personal devices and work devices happens to most employees and it might create a data breaching threat to the business with co-mingling personal and corporate information. The current solutions available to business are network access control, mobile device management, mobile application management, guest management, and content management.

DATA BIASES

Bias is not introduced by AI as new. Biases embedded in the channels of data collection, the approaches of data processing, and the design of statistical modeling. With the increasing popularity of AI, the perils of AI and its algorithmic bias[8] sends the public a significant caveat. However, the human mind is the least transparent logic processing machine. Any past life experience can cast impact on human judgement and lead to bias potentially. On the other hand, it is much easier to investigate and question the processes of AI algorithms. AI unveils the bias problem based on two features of itself: One is that AI processes big stream of data and the impact of bias cumulates in the outputs which is easier to tell than the old calculation techniques; the other is that human bias is more complicated and hard to determine it is a bias.

Around the time GDPR getting effective, May 25, 2018, multiple surveys[9] conducted by different organizations demonstrate the attitude of the general public to balancing their data privacy with the beneficiaries, including personalized services with less biases. In April 2018, Janrain surveyed 1051 American adults showing that 53% are against trading their data for adverts quality improvement. In Marketo's survey data, 36% of the respondents don't put enough trust to businesses keeping and utilizing their personal data. With equally high degree of data privacy issue sensitivity, internet users in Ofcom survey also give a significant disagree ratio on commercializing personal data. Meanwhile, two other institutes get slightly more business-friendly results by emphasizing some customer protection preconditions when using the data. By highlighting personal data will only be used by the first business party rather than third-party vendors, PageFair gets a bigger group of positive feedback, but still 13% would insist on opting out. Likewise, GFK research specifies the use of cookies as the approach to collect personal data and get 19% disagree or somewhat disagree rate.

Organization	Country	Date of Survey	Number of Respondents	Data Privacy Sensitive Rate
Janrain	US	April, 2018	1051	53%
Marketo	UK, FR, and GE	May, 2018	3000	36%
Ofcom	UK	September, 2017	1,875	41%
PageFair	EU	January, 2018	300	13%
GFK research	EU	July, 2017	11020	19%

In the fourth industrial revolution, the data consumption is in a new magnitude with the development of machine learning algorithms, the data scalability and elasticity, and computing power. Flagship companies in telecom industry own data warehouse easily with size of zebibyte. The machine learning models can best fully take advantage of the data with its voracious appetite. Usually, a machine learning model consist of two phases: training and exploitation. These two phases can happen in well distanced space and separated time. Training a model with training data input is similar to educate a five-year-old with examples from a textbook, after finishing the "book", the content of the "book" will be internalized into the "brain" of the machine and it's hard to tell which "neuron" of the "brain" is from the output of training which piece of data. This creates a tricky problem for the industry and data privacy law makers to determine when customer requests for opting that one data point out of the business, if the existing trained model needs to be deleted and retrained with the rest of opt-in data.

On the other hand, if the data point is removed from the training data set, this missing data point creates a hole to the whole research target population. When no enough data to represent features this data point carries, this lack-of-representativeness will convert to biases[10] in every further step and harder to tell the impact of the inherent and ineradicable biases when the machine learning model gets more complicated. This is also why people call this process as "money laundering of biases", since the biases might be blended smoothly into the trend of the data and could be even interpreted as a pseudo-feature of the population.

There are multiple questions emerge when trying to protect both data privacy and the other rights including data bias control: How to protect employee privacy at work while ensuring productivity fulfilling[11]; how to guarantee the interpretability of data usage and explainable machine learning while not hurting the confidentiality of the business; how to build inclusive AI while keeping the detection capabilities of biases; how to encourage responsible sharing and right to repair while preventing data breaching. Overall, it is critical to protect the basic right to choose of individual customer to reveal their true preference, even their choice is not lying in the mainstream.

The laws and regulations should leave cushion space for individuals to trade their personal data for monetary incentives. For example, car insurance companies like Root offer customers usage-based insurance offering discounts if Root can put monitoring devices in their car to track their driving behaviors including driving speed, geographical locations, and other multidimensional data. Other car insurance companies including Progress Insurance have similar driving-habits-based promotion plans. In their plans, it specifically mentions their data privacy solution of the "Snapshot" Privacy Statement[12] with guarantee of data encryption, removing data from local device, and the other data protection Data privacy law should not be used as tools for government to enforce customers ending up with no choices.

No one can highlight less over the importance of data privacy protection. Nevertheless, customer's freedom to choose should also be considered by regulators. In 2016, a survey conducted by Pew Research Center[13] reflects the other angle how customers perceive their data privacy. In this survey, customers explicitly express their willingness of sharing personal information, if the data transferring process is protected and it is directly used for improving their benefits. From this research, business could consider communicating with their customers in a better way by specifying the data categories they collect from customers approaches. Moreover, the company enables policyholders accessible to their data by simply logging in their account.

Maine State government announced one of the most restricted data privacy law[14] to cut off any possibilities for the business to come up with alternative incentives to acquire their customer's personal data. Data privacy law should more focus on helping customers understand how the business ensure their data security and use the data to improve the customer services qualities.

METHODS AND DATA

DATA COLLECTION AND STUDY CASE SELECTION

Breaking the information silos enables the panoramic view of the supply chain of the telecom industry. Synthesizing data from multiple resources both public and private enables consistency of the data by requiring consensus of the data sources majority. Multiple dimensions of unit-level status/performance are recorded with quantifiable values (nominal, ordinal, interval, or ratio) or descriptive English.

Customer data privacy is well protected by dropping data fields directly related to personal identifiable information (PII) and removing any PII data accidentally typed in descriptive columns. The descriptive fields are vectorized with two methods tried separately: bag of words (BoW), and TF-IDF (Term frequency – Inverse Dictionary Frequency). The matrix of the latter can better define the keywords embedded in the context, so TF-IDF was chosen to work in the data preparation step. The impact of the specific triage issue requires the clear scope of the research data pulled in. I.e. we need to understand how the exchange risk performance of certain brand on specific triage issue.

In major carriers' and OEMs' libraries, there are usually 100 to 200 triage codes available for defining the trouble issue(s) of the device. Considering the quantity of these many subgroups and the unbalanced sizes of the triage subgroups, breaking down the complicated multi-classification problem to multiple parallel binary classification problems is a more accurate and efficient solution. Adopting the supervised learning methodology in the comment binary classification step, we need to prepare a training set to train a machine learning model.

Comments are labeled with Boolean values (1 or 0) to respectively represent either it correlated to this triage issue we intent to investigate or uncorrelated. The comments training set will be required with the size of 5k-10k to guarantee there are enough comments in each subgroup and the machine learning model can learn the features of the customer expressions comprehensively.

This model has been practiced with the full historical data of more than ten popular device models in the current market. In this paper, one smartphone flagship device model and its data is chosen as the research target to explain the exchange risk management methodology. Furthermore, we will assess the impact of data bias introduction to data modeling with customers' random opt-out. Eventually, the data will be used to practice the solutions we come up with to decrease the possible impact of the biases and enhance the stability of the statistical model with protecting the model's sensitivity to the potential risk triggers.

VARIABLES SELECTION AND CONTROLLERS

With the data sources guaranteed, we have the multi-dimensions of variables to assess each device unit and track its historical evaluation in the flow of supply chain. IMEI (International Mobile Equipment Identity), a 15 decimal digit serial number is kept as the device unit identifier. Here are the variables which can bring significant impact on the risk fluctuation.

Acquisition date is the Customer Acquisition Event date when this device unit gets acquired. It is recorded in format of MM-DD-YYYY. Chiming in with the format of acquisition date, exchange date is the Customer Relinquish Event date when this device unit gets exchanged. The period of time between the Customer Acquisition Event (CAE) and the Customer Relinquish Event (CRE) is used as days of use to assess how many days the user keeps the phone. Population Growth is the active users of this device model in the market who would consider take actions on their devices including remorse, exchange, or upgrade. This population growth is directly determined by the cumulative number of Customer Acquisition Events (CAE), and Customer Relinquish Events each day.

The firmware version at the date of the Customer Relinquish Event. If no firmware version is available, the MR was determined by the most updated MR available by the Customer Relinquish Event date. The period of time the MR version at exchange is available from upgrade date to the date of the Customer Relinquish Event happening. By assuming the trend of upgrade following uniform distribution, the upgrade date is simulated locating in the range between MR release date and CRE date. Order ID, a nine decimal digit serial number is the order identification of this device unit when getting exchanged. A device unit is assigned with one order ID at each time of exchange. So, it is possible that one device unit can have multiple order ID if it has been exchanged for more than once.

Comments (also named as return remarks) is the description of this device exchange reason provided by the customer at the time of exchange. There are four types of expressions can be found in the comments: problem statement, troubleshooting, inspection, and account information. Problem statement gives the issue that the customer has with the product; Troubleshooting tells the attempt that the user or the customer support agent performs to attempt to fix the issue; Inspection describes the result of the user or customer agent inspecting the product for certain characteristics; Account part provides some necessary customer account information related to this issue. Any PII data in the comments will be removed in the data preparation step before the comments being pulled in the statistical models.

Failure reason ID is an alphanumeric identifier usually consisting of a capital letter and two digits. OEMs, Carriers, third-party repair vendors, and other major stakeholders in the market are using their own system of failure reason ID codes. This also reflects that the system and information silos deeply root in this industry. There is lack of consistency crossing the failure reason ID code systems, including the classification criteria and the count of the subgroups. Most systems would have multiple tiers of classification of the failure reasons like a tree structure: the lower the tier, the more specific the definition of the subgroup. If the system has three tiers, those three tiers can be return type, primary description code, and secondary description code. The lowest tier, the secondary description codes classify the phone issues into 100 to 200 subgroups.

THEORETICAL FRAMEWORK AND HYPOTHESES

The opt-out impact is evaluated with statistical change of exchange risk index model. With the known industry knowledge, MR3 and MR7 are non-successful MR and the other eight MRs are successful ones. We would expect that MR3 and MR7 significantly increase the exchange risk index in their life span.

In the personal data opt out simulation, we assume it follows uniform distribution. The uniform distribution represents customers evenly likely to opt out in the ordinary days. The opt-outs are simulated by evenly removing the data points of customer exchange events in the data collection history. We test the opt-out rate separately as 1%, 2%, 5%, 10%, 20%, 25%, and 30%. Comparing the ERI index fluctuation of the incomplete dataset with the one of the full datasets, interesting analysis results show. The lower data missing rate, the more accurate the ERI would be. This is not surprising. Since the missing data points are evenly spread over the time series, the impact of data holes is not significant in certain time period.

Hypothesis 1: How many days device unit is used by the user will significantly impact the risk .

Hypothesis 2: Exchange risk index can evaluate the performance of MRs, i.e. the successful MRs would decrease the exchange risk, while the non-successful MRs would destabilize the exchange risk and will not effectively decrease the risk.

Hypothesis 3: With personal data opt out, it might affect the model to tell the MR performance by telling whether it can significantly lower the risk.

Hypothesis 4: The dataset with opt-outs will affect the statistical model quality. The exchange risk will decrease which does not honestly reflect the real risk and causes the inaccuracy of ERI.

DEPENDENT SELECTION

To lower the count of daily exchange and eventually to control the A-stock device value drop and its related post-sale reverse logistics cost is the end goal of the telecom industry. However, it is not a good idea to set the daily exchange count as the only metric to evaluate the performance of the device model. Customer exchange behavior is a very individualized decision-making process. Different triggers might have different degree of impact on each customer. For example, some customers cannot tolerate a phone with cracked screen, but some others feel comfortable with the cracked screen if it does not affect the phone performing the major functions, like calling or texting. This phenomenon also decreases the predictability of the naïve metric of daily exchange count. Another reason is about the comparability of this metric. It is unfair to compare the performance of different device models on the same trouble issue or to prioritize the trouble issue to solve happening to the same device model by using daily exchange count. For example, the popular device models with large population can lead to high daily exchange count even with high quality device. The daily exchange of certain trouble issue will be normalized by the active user population of this device model and the life span of its MR version at exchange. Here MR availability in the market is also a critical factor to control. After the new MR release, for the users who would choose to update it, it usually takes three days to one week or even longer to finish their OS update. Considering that different MRs have different life spans, the longer the MR stays available in the market the more likely the customer would have chance to try it. If a MR life span is short, the customer might not have enough time to find its problem, or directly jump to the next MR.

RESULTS

In this smartphone flagship device model data set, it consists of 7927 data records of device exchanges. Table 1 reports descriptive statistics and correlations for study variables relevant to the exchange risk. The overall average days users keeping their phones with them (days of use) is 144.7 days, and its standard deviation as 114.03 (M = 144.7, SD = 114.03). This number indicates the users would still give enough time to test the phones before they decide to return it. Meanwhile, the variance of days of use is high. Some users would change their mind to return their phone very soon after their purchase, which might be due to their taking advantage of the remorse rule (within 14 days full price refund no reason needed). The average daily exchange is also a very hard to predict variable (M = 21.26, SD = 8.19). Averagely, an operation system version, MR (M = 32.78, SD = 31.99) would last in the market for 30 to 50 days. In our data, there are 25% of exchanges happen within 11 days of upgrading to the new MR and the majority (53%) happen within 26 days. It shows the software instability issue do have strong correlation to their MR upgrade, which might directly cause the incompatible between the operation system and all the apps in user’s phone.

Table 1 Exchange Risk Data Descriptive Statistics and Correlations

	Days of use	Daily exchange	Cumulative total population	MR availability
Mean	144.7	21.26	189474.02	32.78
STD	114.03	8.19	66843.01	31.99
Daily exchange	0.303***			
Cumulative total population	0.540***	0.543***		
MR availability	0.192***	0.100***	0.270***	

Table 2 contains results from the generalized linear models testing Hypotheses 1 to 3. Exchange risk index can evaluate the performance of MRs: The MR dummy variables coefficients correspond to the industry MR evaluation, which gives MR3 and MR7 would increase the exchange risk, and the other MRs can significantly decrease the risk. However, when 20% of the data opt out, the predictor significance change. The after opt-out coefficient of MR4 shows it does not significantly impact the exchange risk. Meanwhile, even the adjusted R-square does not change, the model F-statistic decreases.

Table 2 Standardized Regression Coefficients Predicting Exchange Risk Index

	ERI model coefficient	
	With full dataset	With 20% data opt out
constant	0.0004***	0.0004
Days of use	-1.44E-07**	-1.67E-07**
mr1	-0.0003***	-0.0003***
mr2	-0.0002***	-0.0002***
mr3	0.0004***	0.0004***
mr4	-4.42E-05**	-4.00E-05
mr5	-0.0002***	-0.0002***
mr6	-0.0003***	-0.0003***
mr7	0.0016***	0.0017***
mr8	-0.0003***	-0.0003***
mr9	-0.0001***	-0.0001***
mr10	-0.0001***	-0.0001***
Overall Model F	238.1***	201.7***
Adjusted R2	0.23	0.23
Note. A positive coefficient indicates the risk elevation in the MR.		
*p < .05; **p < .01; ***p < .001.		

Hypothesis 4 is to test whether there is a statistically significant difference over time between the two ERIs. The T test results by comparing the difference between any ERI with opt-out data and ERI with full data $y_{diff}=y_{full}-y_{missing}$ with null hypothesis. With two-tailed p-value computed using the t distribution. It is the probability of observing a greater absolute value of t under the null hypothesis. With the pre-specified alpha level as 0.05, the two-tail P(T<=t) for the 1% missing data is 0.03, 5% missing data 0.256, 10% missing data 5.56E-07, and 30% missing data 0.25. Comparing the pre-specified alpha level with two-tail P-value, we can conclude that the difference gap between output with full dataset and output with random missing value is not always statistically insignificantly different from zero. This result shows that even opting out individual exchange data points follows uniform distribution, the ERI index as the output of the incomplete dataset will be off the accurate values. And the incomplete dataset output mean might also be significantly different from the one of the complete datasets.

DISCUSSION

This paper is an attempt to explore the connection between two customer benefits centered topics: data privacy protection and decreasing the data biases. With the impact of trendy data privacy news and policies, customers come up with new perceptions on the balance between data privacy and the convenience from their personalized services. However, overcorrecting the data privacy issue by shrinking customer's pool of choices is hurting customer's rights. In telecom industry, the business should consider initiating self-regulations by improving the data security and communicating their data utilization plan to their customer, the individual data owner.

On the other hand, known the adopting AI and statistical modeling to support business decisions, telecom industry is encouraged to exploit its rich legacy data storage and keep comprehensive data records in the future. This paper shows the significantly difference caused by the missing data shown in this analysis and the possible damage this stats inaccuracy can affect business's market interpretation. Business should recognize this caveat and consider introducing solutions to more strict governmental policies. We can expect more data opt-outs can happen with unpredictable trend. Proper statistical solutions to the data holes will prevent business from introducing biases to their decision-making process.

REFERENCE

[1]Wachter, Sandra. "Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR." Computer law & security review 34.3 (2018): 436-449.
[2]"Telecommunication." Wikipedia, Wikimedia Foundation, 17 Aug. 2019, https://en.wikipedia.org/wiki/Telecommunication.
[3Bloomberg].com, Bloomberg, https://www.bloomberg.com/news/articles/2019-07-12/ftc-approves-facebook-privacy-settlement-worth-about-5-billion.
[4]Karjoth, Günter, Matthias Schunter, and Michael Waidner. "Platform for enterprise privacy practices: Privacy-enabled management of customer data." International Workshop on Privacy Enhancing Technologies. Springer, Berlin, Heidelberg, 2002.
[5]Mosendz, Polly. "Android's Factory Reset Doesn't Delete Everything. Here's How to Really Wipe Your Data." The Atlantic, Atlantic Media Company, 10 July 2014,

https://www.theatlantic.com/technology/archive/2014/07/android-factory-reset-doesnt-delete-everything-heres-how-to-really-wipe-your-data/374192/.

[6] Schwamm, Riqui, and Neil C. Rowe. "Effects of the factory reset on mobile devices." Journal of Digital Forensics, Security and Law 9.2 (2014): 17.

[7] Storm, Darlene, and Darlene Storm. “Think You Deleted Your Dirty Little Secrets? Before You Sell Your Android Smartphone...” Computerworld, Computerworld, 9 July 2014, https://www.computerworld.com/article/2476496/think-you-deleted-your-dirty-little-secrets-before-you-sell-your-android-smartphone.html.

[8] Miller, Alex P. “Want Less-Biased Decisions? Use Algorithms.” Harvard Business Review, 26 July 2018, https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms?referral=03759&cm_vc=rr_item_page.bottom.

[9] AdExchanger // Friday, July 6th. “Post-GDPR, How Many Will Really Opt Out Of Personal Targeting?” AdExchanger, 5 July 2018, https://adexchanger.com/data-driven-thinking/post-gdpr-how-many-will-really-opt-out-of-personal-targeting/.

[10] “Privacy Rights and Data Collection in a Digital Economy: United States Committee on Banking, Housing, and Urban Affairs.” Hearing | Hearings | United States Committee on Banking, Housing, and Urban Affairs, https://www.banking.senate.gov/hearings/privacy-rights-and-data-collection-in-a-digital-economy.

[11] Bélanger, France, and Robert E. Crossler. "Privacy in the digital age: a review of information privacy research in information systems." MIS quarterly 35.4 (2011): 1017-1042.

[12] “Snapshot® Privacy Statement.” Progressive, https://www.progressive.com/support/legal/snapshot-privacy-statement/.

[13] Rainie, Lee, and Maeve Duggan. “Auto Trackers Not Worth Car Insurance Discounts, Most Say.” Pew Research Center: Internet, Science & Tech, 15 Jan. 2016, https://www.pewinternet.org/2016/01/14/scenario-auto-insurance-discounts-and-monitoring/.

[14] “Maine Follows California Lead: Prohibits ISP Use, Sale, Disclosure of Online Consumer Information Without Prior Affirmative Consent.” The National Law Review, https://www.natlawreview.com/article/maine-follows-california-lead-prohibits-isp-use-sale-disclosure-online-consumer.