A Tool for Inclusion or Exclusion?

Contents

Executive Summary . . . . . **i**

I. Introduction **1**

II. Life Cycle of Big Data **3**

III. Big Data's Benefits and Risks **5**

IV. Considerations for Companies in Using Big Data **12**

A. Potentially Applicable Laws 12

Questions for Legal Compliance 24

B. Special Policy Considerations Raised by Big Data Research 25

Summary of Research Considerations 32

V. Conclusion **33**

Appendix:

Separate Statement of Commissioner Maureen K. Ohlhausen **A-1**

Executive Summary

We are in the era of big data. With a smartphone now in nearly every pocket, a computer in nearly every household, and an ever-increasing number of Internet-connected devices in the marketplace, the amount of consumer data owing throughout the economy continues to increase rapidly.

e analysis of this data is often valuable to companies and to consumers, as it can guide the development of new products and services, predict the preferences of individuals, help tailor services and opportunities, and guide individualized marketing. At the same time, advocates, academics, and others have raised concerns about whether certain uses of big data analytics may harm consumers, particularly low-income and underserved populations.

To explore these issues, the Federal Trade Commission ("FTC" or "the Commission") held a public workshop, *Big Data: A Tool for Inclusion or Exclusion?*, on September 15, 2014. e workshop brought together stakeholders to discuss both the potential of big data to create opportunities for consumers and to exclude them from such opportunities. e Commission has synthesized the information from the workshop, a prior FTC seminar on alternative scoring products, and recent research to create this report.

ough "big data" encompasses a wide range of analytics, this report addresses only the commercial use of big data consisting of consumer information and focuses on the impact of big data on low-income and underserved populations. Of course, big data also raises a host of other important policy issues, such as notice, choice, and security, among others. ose, however, are not the primary focus of this report.

As "little" data becomes "big" data, it goes through several phases. e life cycle of big data can be divided into four phases: (1) collection; (2) compilation and consolidation; (3) analysis; and (4) use.

is report focuses on the fourth phase and discusses the bene ts and risks created by the use of big data analytics; the consumer protection and equal opportunity laws that currently apply to big data; research in the eld of big data; and lessons that companies should take from the research. Ultimately, this report is intended to educate businesses on important laws and research that are relevant to big data analytics and provide suggestions aimed at maximizing the bene ts and minimizing its risks.

Big Data's Benefits and Risks

Big data analytics can provide numerous opportunities for improvements in society. In addition to more e ectively matching products and services to consumers, big data can create opportunities for low-

Consumer Protection Laws Applicable to Big Data

Workshop participants and commenters discussed how companies can use big data in ways that provide bene ts to themselves and society, while minimizing legal and ethical risks. Speci cally, they noted that companies should have an understanding of the various laws, including the Fair Credit Reporting Act, equal

Ultimately, as with the FCRA, whether a practice is unlawful under equal opportunity laws is a case-speci c inquiry, and as such, companies should proceed with caution when their practices could result in disparate treatment or have a demonstrable disparate impact based on protected characteristics.

3. The Federal Trade Commission Act

Workshop participants and commenters also discussed the applicability of Section 5 of the Federal Trade Commission Act ("FTC Act"), which prohibits unfair or deceptive acts or practices, to big data analytics. Companies engaging in big data analytics should consider whether they are violating any material promises to consumers—whether that promise is to refrain from sharing data with third parties, to provide consumers with choices about sharing, or to safeguard consumers' personal information—or whether they have failed to disclose material information to consumers. In addition, companies that maintain big data on consumers should take care to reasonably secure consumers' data. Further, at a minimum, companies must not sell their big data analytics products to customers if they know or have reason to know that those customers will use the products for fraudulent or discriminatory purposes. e inquiry will be fact-speci c, and in every case, the test will be whether the company is o ering or using big data analytics in a deceptive or unfair way.

Research on Big Data

Workshop participants, academics, and others also addressed the ways big data analytics could a ect low-income, underserved populations, and protected groups. Some pointed to research that demonstrates that there is a potential for incorporating errors and biases at every stage—from choosing the data set used to make predictions, to de ning the problem to be addressed through big data, to making decisions based on the results of big data analysis—which could lead to potential discriminatory harms. Others noted that these concerns are overstated or simply not new, and emphasized that rather than disadvantaging minorities, big data can create opportunities for low-income and underserved populations.

To maximize the bene ts and limit the harms of big data, the Commission encourages companies to consider the following questions raised by research in this area:

How representative is your data set? Companies should consider whether their data sets are missing information about certain populations, and take steps to address issues of underrepresentation and overrepresentation. For example, if a company targets services to consumers who communicate through an application or social media, they may be neglecting populations that are not as tech-savvy.

Does your data model account for biases? Companies should consider whether biases are being incorporated at both the collection and analytics stages of big data's life cycle, and develop strategies to overcome them. For example, if a company has a big data algorithm that only considers applicants from "top tier" colleges to help them make hiring decisions, they may be incorporating previous biases in college admission decisions.

How accurate are your predictions based on big data? Companies should remember that while big data is very good at detecting correlations, it does not explain which correlations are meaningful. A prime example that demonstrates the limitations of big data analytics is Google Flu Trends, a machine-learning algorithm for predicting the number of u cases based on Google search terms. While, at rst, the algorithms appeared to create accurate predictions of where the u was more prevalent, it generated highly inaccurate estimates over time. is could be because the algorithm failed to take into account certain variables. For example, the algorithm may not have taken into account that people would be more likely to search for u-related terms if the local news ran a story on a u outbreak, even if the outbreak occurred halfway around the world.

Does your reliance on big data raise ethical or fairness concerns? Companies should assess the factors that go into an analytics model and balance the predictive value of the model with fairness considerations. For example, one company determined that employees who live closer to their jobs stay at these jobs longer than those who live farther away. However, another company decided to exclude this factor from its hiring algorithm because of concerns about racial discrimination, particularly since di erent neighborhoods can have di erent racial compositions.

e Commission encourages companies to apply big data analytics in ways that provide bene ts and opportunities to consumers, while avoiding pitfalls that may violate consumer protection or equal opportunity laws, or detract from core values of inclusion and fairness. For its part, the Commission will continue to monitor areas where big data practices could violate existing laws, including the FTC Act, the FCRA, and ECOA, and will bring enforcement actions where appropriate. e Commission will also continue to examine and raise awareness about big data practices that could have a detrimental impact on low-income and underserved populations, and promote the use of big data that has a positive impact on such populations.

I. Introduction

e era of big data has arrived. While companies historically have collected and used information about their customer interactions to help improve their operations, the expanding use of online technologies has greatly increased the amount of consumer data that ows throughout the economy. In many cases, when consumers engage digitally—whether by shopping, visiting websites, paying bills, connecting with family and friends through social media, using mobile applications, or using connected devices, such as tness trackers or smart televisions—companies collect information about their choices, experiences, and individual characteristics. e analysis of this consumer information is often valuable to companies and to consumers, as it provides insights into market-wide tastes and emerging trends, which can guide the development of new products and services. It is also valuable to predict the preferences of speci c individuals, help tailor services, and guide individualized marketing of products and services.

e term "big data" refers to a con uence of factors, including the nearly ubiquitous collection of consumer data from a variety of sources, the plummeting cost of data storage, and powerful new capabilities to analyze data to draw connections and make inferences and predictions.[1]

A common framework for characterizing big data relies on the "three Vs," the volume, velocity, and variety of data, each of which is growing at a rapid rate as technological advances permit the analysis and use of this data in ways that were not possible previously.[2] Volume refers to the vast quantity of data that can be gathered and analyzed e ectively. e costs of collecting and storing data continue to drop dramatically. And the ability to access millions of data points increases the predictive power of consumer data analysis.

1 *See, e.g.*, Exec. Office of the President, Big Data: Seizing Opportunities, Preserving Values 2–3 (2014) [hereinafter "White House May 2014 Report"], http://www.whitehouse.gov/sites/default/ les/docs/big_data_privacy_report_5.1.14_ nal_print.pdf; Jim atcher, *Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data*, 8 Int'l J. Of Commc'n 1765, 1767–69 (2014), http://ijoc.org/index.php/ijoc/article/view/2174/1158. *See also* Comment #00018 from Persis Yu, Nat'l Consumer L. Ctr., to Fed. Trade Comm'n, attached report at 10 (Aug. 15, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/08/00018-92374.pdf.

2 *See, e.g.*, Transcript of Big Data: A Tool for Inclusion or Exclusion?, in Washington, D.C. (Sept. 15, 2014), at 15 (Solon Barocas), 32 (Joseph Turow), 40–41 (Joseph Turow), 261 (Christopher Wolf) [hereinafter Big Data Tr.], https://www.ftc.gov/system/ les/documents/public_events/313371/bigdata-transcript-9_15_14.pdf. *See also* White House May 2014 Report, *supra* note 1, at 4–5; Comment #00067 from Mark MacCarthy, Software & Info. Indus. Assoc., to Fed. Trade Comm'n 2 (Oct. 31, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00067-92918.pdf; Comment #00065 from Jules Polonetsky & Christopher Wolf, Future of Privacy Forum, to Fed. Trade Comm'n 2 (Oct. 31, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00065-92921.pdf; Comment #00049 from Martin Abrams, Info. Accountability Found., to Fed. Trade Comm'n 3–4 & n.6, https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00049-92780.pdf; Comment #00031 from M. Gary LaFever & Ted Myerson, anonos, to Fed. Trade Comm'n 1 (Aug. 21, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/08/00031-92442.pdf. Others suggest that there is a "fourth V," veracity, to denote the accuracy and integrity of data used. *See, e.g.*, Brian Gentile, *e New Factors of Production and the Rise of Data-Driven Applications*, Forbes (Oct. 31, 2011), http://www.forbes.com/sites/ciocentral/2011/10/31/the-new-factors-of-production-and-the-rise-of-data-driven-applications/.

Velocity is the speed with which companies can accumulate, analyze, and use new data. Technological improvements allow companies to harness the predictive power of data more quickly than ever before, sometimes instantaneously.[3]

Variety means the breadth of data that companies can analyze e ectively. Companies can now combine very di erent, once unlinked, kinds of data—either on their own or through data brokers or analytics rms—to infer consumer preferences and predict consumer behavior, for example.

Together, the three Vs allow for more robust research and correlation. Previously, nding a representative data sample su cient to produce statistically signi cant results could be very di cult and expensive. Today, the present scope and scale of data collection enables cost-e ective, substantial research of even obscure or mundane topics (e.g., the amount of foot tra c in a park at di erent times of day).

Big data can produce tremendous bene ts for society, such as advances in medicine, education, health, and transportation, and in many instances, without using consumers' personally identi able information. Big data also can allow companies to improve their o erings, provide consumers with personalized goods and services, and match consumers with products they are likely to be interested in. At the same time, advocates, academics, and others have raised concerns about whether certain uses of big data analytics may harm consumers. For example, if big data analytics incorrectly predicts that particular consumers are not likely to respond to prime credit o ers, certain types of educational opportunities, or job openings requiring a college degree, companies may miss a chance to reach individuals that desire this information. In addition, if big data analytics incorrectly predicts that particular consumers are not good candidates for prime credit o ers, educational opportunities, or certain lucrative jobs, such educational opportunities, employment, and credit may never be o ered to these consumers. Some fear that such incorrect predictions could perpetuate existing disparities.

To examine these issues, the Federal Trade Commission ("FTC" or "the Commission") held a public workshop, *Big Data: A Tool for Inclusion or Exclusion?*, on September 15, 2014.[4] In particular, the workshop explored the potential impact of big data on low-income and underserved populations. e workshop brought together academics, government representatives, consumer advocates, industry representatives, legal practitioners, and others to discuss the potential of big data to create opportunities for consumers or exclude them from such opportunities. e workshop consisted of four panels addressing the following topics: (1) current uses of big data; (2) potential uses of big data; (3) the application of equal opportunity and consumer protection laws to big data; and (4) best practices to enhance consumer protection in the use of big data. e Commission also received sixty- ve public comments on these issues from private citizens, industry representatives, trade groups, consumer and privacy advocates, think tanks, and academics.

3 White House May 2014 Report, *supra* note 1, at 5.

4 e materials from the workshop are available on the FTC website at http://www.ftc.gov/news-events/events-calendar/2014/09/big-data-tool-inclusion-or-exclusion.

e Commission has synthesized the discussions and comments from the workshop—along with the record from a prior FTC seminar on alternative scoring products[5] and recent research—to create this report, which focuses on the impact of big data on low-income and underserved populations. e report is divided into four sections. First, the report describes the "life cycle" of big data and how "little" data turns into big data. Second, it discusses some of the bene ts and risks created by the use of big data. ird, it describes some of the consumer protection laws that currently apply to big data. Finally, it discusses certain research in the eld of big data and lessons that companies should take from the research in order to help them maximize the bene ts of big data while mitigating risks. Importantly, though the term "big data" encompasses a wide range of analytics, this report addresses only the commercial use of big data consisting of consumer information.[6]

II. Life Cycle of Big Data

e life cycle of big data can be divided into four phases: (1) collection; (2) compilation and consolidation; (3) data mining and analytics; and (4) use.[7]

As to the rst step, not all data starts as big data. Rather, companies collect bits of data from a variety of sources.[8] For example, as consumers browse the web or shop online, companies can track and link their activities. Sometimes consumers log into services or identify themselves when they make a purchase. Other

5 On March 19, 2014, the Commission hosted a seminar on alternative scoring products and received nine public comments in connection with the seminar. *Spring Privacy Series: Alternative Scoring Products*, Fed. Trade Comm'n (Mar. 19, 2014), http://www.ftc.gov/news-events/events-calendar/2014/03/spring-privacy-series-alternative-scoring-products.

6 e report does include some examples from non-commercial elds, but it is intended to guide companies as they use big data about consumers.

7 *See, e.g.*, Nat'l Consumer L. Ctr. Comment #00018, *supra* note 1, attached report at 11–12. In May 2014, the Commission released a report studying data brokers, which focused on the rst three phases of the life cycle of big data. Fed. Trade

of statistical models to generate new data.[18] Developing and testing the models that nd patterns and make predictions can require the collection and use of copious amounts of data.[19] In a market context, a common purpose of big data analytics is to draw inferences about consumers' likely choices. Companies may decide to adopt big data analytics to better understand consumers, potentially by using data to attribute to an individual the qualities of those who appear statistically similar, e.g., those who have made similar decisions in similar situations in the past. us, a retail rm might use data on its customers' past purchases, web searches, shopping habits, and prices paid to create a statistical model of consumers' purchases at di erent prices. With that model, the retailer could then compare a prospective consumer's characteristics or past purchases, web searches, and location information to predict how likely the consumer is to purchase a product at various price points.

e nal step in the life cycle of big data is use. e Commission's May 2014 report entitled *Data Brokers: A Call for Transparency and Accountability* focused on the rst three steps in the life cycle of big data within that industry—collection, compilation, and analytics.[20] It discussed how information gathered for one purpose (e.g., paying for goods and services) could be compiled and analyzed for other purposes, such as for marketing or risk mitigation. In contrast, this report focuses on certain *uses* of big data. It examines the question of how companies use big data to help consumers and the steps they can take to avoid inadvertently harming consumers through big data analytics.

III. Big Data's Benefits and Risks

Companies have been analyzing data from their own customer interactions on a smaller scale for many years, but the era of big data is still in its infancy.[21] As a result, mining large data sets to nd useful, non-obvious patterns is a relatively new but growing practice in marketing, fraud prevention, human resources, and a variety of other elds. Companies are still learning how to deal with big data and unlock its potential while avoiding unintended or unforeseen consequences.[22]

Appropriately employing big data algorithms on data of su cient quality can provide numerous opportunities for improvements in society. In addition to the market-wide bene ts of more e ciently matching products and services to consumers, big data can create opportunities for low-income and

18 *See, e.g.*, Galit Shmueli, *To Explain or Predict?*, 25 Statistical TfuTolu/Span <MCID 615 >BDC<MCID 623 >BDC uBTu/T1_1 1 Tfu12 0 0

underserved communities.[23] Workshop participants and others have noted that big data is already being used to:

Increase educational attainment for individual students. Educational institutions have used big data techniques to identify students for advanced classes who would otherwise not have been eligible for such classes based on teacher recommendations alone.[24] ese institutions have also used big data techniques to help identify students who are at risk of dropping out and in need of early intervention strategies.[25] Similarly, organizations have used big data analytics to demonstrate how certain disciplinary practices, such as school suspensions, a ect African-American students far more than Caucasian students, thereby partly explaining the large discrepancy between the graduation rates of these two groups.[26]

Provide access to credit using non-traditional methods. Companies have used big data to provide alternative ways to score populations that were previously deemed unscorable.[27] For example, LexisNexis has created an alternative credit score called RiskView.[28] is product relies on traditional public record information, such as foreclosures and bankruptcies, but it also includes educational history, professional licensure data, and personal property ownership data. us, consumers who may not have access to traditional credit, but, for instance, have a professional license, pay rent on time, or own a car, may be given better access to credit than they otherwise would have.[29]

23 *See, e.g.*, Big Data Tr. 83–85 (Mark MacCarthy), 250–51 (Christopher Wolf). *See generally* Comment #00076 from William Kovacs, U.S. Chamber of Commerce, to Fed. Trade Comm'n (Oct. 31, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00076-92936.pdf; Comment #00073 from Michael Beckerman, e Internet Assoc., to Fed. Trade Comm'n (Oct. 31, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00073-92923.pdf; Comment #00066 from Carl Szabo, NetChoice, to Fed. Trade Comm'n (Oct. 31, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00066-92920.pdf; Comment #00063 from Peggy Hudson, Direct Mktg. Assoc., to

Furthermore, big data algorithms could help reveal underlying disparities in traditional credit markets and help companies serve creditworthy consumers from any background.[30]

Provide healthcare tailored to individual patients' characteristics. Organizations have used big data to predict life expectancy, genetic predisposition to disease, likelihood of hospital readmission, and likelihood of adherence to a treatment plan in order to tailor medical treatment to an individual's characteristics.[31] is, in turn, has helped healthcare providers avoid one-size- ts-all treatments and lower overall healthcare costs by reducing readmissions.[32] Ultimately, data sets with richer and more complete data should allow medical practitioners more e ectively to perform "precision medicine," an approach for disease treatment and prevention that considers individual variability in genes, environment, and lifestyle.[33]

Provide specialized healthcare to underserved communities. IBM, for example, has worked with hospitals to develop an Oncology Diagnosis and Treatment Advisor. is system synthesizes vast amounts of data from textbooks, guidelines, journal articles, and clinical trials to help physicians make diagnoses and identify treatment options for cancer patients. In rural and low-income areas, where there is a shortage of specialty providers, IBM's Oncology Diagnosis and Treatment Advisor can provide underserved communities with better access to cancer care and lower costs.[34]

Increase equal access to employment. Companies have used big data to help promote a more diverse workforce.[35] Google, for example, recognized that its traditional hiring process was resulting in a homogenous work force. rough analytics, Google identi ed issues with its hiring process, which included an emphasis on academic grade point averages and "brainteaser" questions

30 *See, e.g.*, C248.4 Data-Driven Innovation for Growth and Well-Being, Interim Synthesis Report, OECD, 8 (2014), http://www.oecd.org/sti/inno/data-driven-innovation-interim-synthesis.pdf; Big Data: Seizing Opportunities, Preserving Values, Exec. Office of the President, 53 (2014), https://www.whitehouse.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Report.pdf; Big Data: A Tool for Inclusion or Exclusion?, Federal Trade Commission, 3 (2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf; Big Data and Innovation, Key Themes for Future Economic Growth, OECD, 20 (2015), http://www.oecd.org/sti/inno/Big-Data-Innovation-Key-Themes-for-Future-Economic-Growth

during interviews. Google then modi ed its interview practices and began asking more structured behavioral questions (e.g., how would you handle the following situation?).[36] is new approach helped ensure that potential interviewer biases had less e ect on hiring decisions.

While recognizing these potential bene ts, some researchers and others have expressed concern that the use of big data analytics to make predictions may exclude certain populations from the bene ts society and markets have to o er. is concern takes several forms. First, some workshop participants and commenters expressed concerns about the [q]uality of data, including its accuracy, completeness, and representativeness.[37] Some academic research has found that hidden biases in the [data] sets used to train predictive algorithms can lead to unfair outcomes.[38] Second, some workshop participants and commenters raised concerns about the potential for big data analytics to result in discrimination against protected classes.[39] Some workshop participants and commenters also raised concerns about the potential for big data analytics to result in discrimination against low-income consumers.[40] Finally, some workshop participants and commenters raised concerns about the potential for big data analytics to result in discrimination against consumers in ethnic communities.[41]

Second, while big data may be highly e ective in showing correlations, it is axiomatic that correlation is not causation.[42] Indeed, with large enough data sets, one can generally nd some meaningless correlations. For example, in eighteen out of the past twenty U.S. Presidential elections, if the Washington, D.C. professional football team won its last home game before the election, the incumbent's party continued to hold the presidency; if the team lost that last home game, the out-of-o ce party unseated the incumbent party.[43] Other examples of spurious correlations abound.[44] If companies use correlations to make decisions about people without understanding the underlying reasons for the correlations, those decisions might be faulty and could lead to unintended consequences or harm for consumers and companies.

Ultimately, all of these concerns feed into the larger concern about whether big data may be used to categorize consumers in ways that can result in exclusion of certain populations. Workshop participants and others have noted how potential inaccuracies and biases might lead to detrimental e ects for low-income and underserved populations.[45] According to these commenters, particular uses of big data may:

> **Result in more individuals mistakenly being denied opportunities based on the actions of others.** Participants raised concerns that big data can lead to decision-making based on the actions of others with whom consumers share some characteristics.[46] Several commenters explained that some credit card companies have lowered a customer's credit limit, not based on the customer's payment history, but rather based on analysis of other customers with a poor repayment history that had shopped at the same establishments where the customer had shopped.[47] Indeed, one credit card company settled FTC allegations that it failed to disclose its practice of rating consumers as having a greater credit risk because they used their cards to pay for marriage counseling, therapy, or tire-repair services, based on its experiences with other consumers and their repayment histories.[48] Using this type of a statistical model might reduce the cost of credit for some individuals, but may also result

42 *See generally* John Aldrich, *Correlations Genuine and Spurious in Pearson and Yule*, 10(4) Statistical Sci. 364 (1995), http://www.jstor.org/stable/2246135. *See also Correlation*, xkcd, https://xkcd.com/552/ (last visited Dec. 29, 2015).

43 *Winning Tradition*, Snopes.com, http://www.snopes.com/politics/ballot/redskins.asp (last visited Dec. 29, 2015).

44 *See, e.g.*, *Spurious Correlations*, Tylervigen.com, http://www.tylervigen.com/spurious-correlations (last visited Dec. 29, 2015) (showing a variety of spurious correlations, including, for example, a historical correlation between the annual number of people who have drowned by falling into a swimming pool and the annual number of lms in which Nicolas Cage has appeared).

45 *See, e.g.*, Big Data Tr. 222 (Jeremy Gillula). *See generally* Nat'l Consumer L. Ctr. Comment #00018, *supra* note 1; Common Sense Media Comment #00016, *supra* note 8; N.Y.U. Info. L. Inst. Comment #00015, *supra* note 8; Ctr. for Dig. Democracy & U.S. PIRG Educ. Fund Comment #00003, *supra* note 8.

46 *See, e.g.*, Big Data Tr. 42–44 (danah boyd). *See also* Comment #00078 from Seeta Peña Gangadharan et al., New Am.'s Open Tech. Inst., to Fed. Trade Comm'n, attached report entitled, The Networked Nature of Algorithmic Discrimination, at 53–57 (Oct. 31, 2014), https://www.ftc.gov/system/ les/documents/public_comments/2014/10/00078-92938.pdf.

47 *See, e.g.*, Nat'l Consumer L. Ctr. Comment #00018, *supra* note 1, at 27–28; N.Y.U. Info. L. Inst. Comment #00015, *supra* note 8, at 6.

48 *See* FTC v. CompuCredit Corp., No. 1:08-cv-1976-BBM-RGV (N.D. Ga. June 10, 2008), https://www.ftc.gov/sites/default/ les/documents/cases/2008/12/081219compucreditstiporder.pdf. *See also* Danielle Keats Citron & Frank A. Pasquale III, *e Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1, 4 (2014), http://ssrn.com/abstract=2376209.

in some creditworthy consumers being denied or charged more for credit than they might otherwise have been charged.[49]

Create or reinforce existing disparities. Participants raised concerns that when big data is used to target ads, particularly for nancial products, low-income consumers who may otherwise be eligible for better o ers may never receive them.[50]

Expose sensitive information. Participants also raised concerns about the potential exposure of characteristics that people may view as sensitive.[51] For example, one study combined data on Facebook "Likes" and limited survey information to determine that researchers could accurately predict a male user's sexual orientation 88 percent of the time; a user's ethnic origin 95 percent of time; and whether a user was Christian or Muslim (82 percent), a Democrat or Republican (85 percent), or used alcohol, drugs, or cigarettes (between 65 percent and 75 percent).[52]

Assist in the targeting of vulnerable consumers for fraud. Unscrupulous companies can use big data to o er misleading o ers or scams to the most vulnerable prospects.[53] According to public reports, unscrupulous companies can obtain lists of people who reply to sweepstakes o ers and thus are more likely to respond to enticements, as well as lists of "su ering seniors" who are identi ed as having Alzheimer's or similar maladies.[54] Big data analytics allows companies to more easily and accurately identify such vulnerable prospects.

Create new justi cations for exclusion. Big data analytics may give companies new ways to attempt to justify their exclusion of certain populations from particular opportunities. For example, one big data analytics study showed that "people who ll out online job applications using browsers that did not come with the computer . . . but had to be deliberately installed (like Firefox or Google's

49 *See, e.g.*

As these examples show, big data o ers companies the opportunity to facilitate inclusion or exclusion. Companies can use big data to advance education, credit, and employment opportunities for low-income communities or to exclude them from these opportunities. ey can use big data to target products to those who are most interested or to target products in ways that could exclude certain populations. e remainder of this report is intended to guide companies on some of the laws that may apply when using big data, raise awareness about the ethical implications of using big data, and to highlight potential biases that companies should consider as they use big data.

IV. Considerations for Companies in Using Big Data

e challenge for companies is not *whether* they should use big data; indeed, the reality of today's marketplace is that big data now fuels the creation of innovative products and systems that consumers and companies quickly are coming to rely upon and expect. Rather, the challenge is *how* companies can use big data in a way that bene ts them and society, while minimizing legal and ethical risks.

In assessing risks, companies should rst have an understanding of the laws that may apply to big data practices. Second, they should be aware of important research in the eld of big data aimed at identifying potential biases and inaccuracies. D 14

departments in making hiring decisions.[69] Based on the allegations that the company marketed consumer pro les speci cally for employment purposes, the Commission charged that Spokeo was subject to, but had failed to comply with, the FCRA. Accordingly, the FTC entered into a consent decree that required Spokeo to pay $800,000 in civil penalties.

In another matter, the Commission alleged that the data broker Instant Checkmate advertised potential uses of its consumer data for employment and tenant screening purposes, both through its website and through blog posts, but did not comply with the FCRA.[70] According to the complaint, the company used a Google AdWords campaign to display ads for its services that would appear in search results when consumers sought background checks on "nannies," "babysitters," "maids," and "housekeepers." us, the Commission alleged that the company was subject to the FCRA, entered into a consent order to ensure future compliance, and obtained $550,000 in civil penalties.[71] In both *Spokeo* and *Instant Checkmate*, the companies included a disclaimer on their websites stating that they were not CRAs and that users could not use their data for eligibility purposes. ese disclaimers were not e ective in insulating the companies from FTC enforcement. As these cases demonstrate, the scope of the FCRA extends beyond traditional credit bureaus.

Companies that use consumer reports also have obligations under the FCRA. ey must, among other things, provide consumers with "adverse action" notices if the companies use the consumer report information to deny credit, insurance, employment, housing, or certain other covered bene ts.[72] Similarly, companies that use consumer reports must provide "risk-based pricing" notices if they charge consumers more to obtain credit or insurance based on consumer report information.[73] e purpose of both types of notices is to enable consumers to check their consumer reports and correct any inaccuracies.[74] e Commission has brought actions against various companies for violation of these provisions.[75] For example,

69 Complaint at 3–4, *Spokeo*, No. 2-12-cv-05001-MMM-SH (C.D. Cal. led June 7, 2012), https://www.ftc.gov/sites/default/ les/documents/cases/2012/06/120612spokeocmpt.pdf.

70 Complaint, United States v. Instant Checkmate, Inc., No. 3:14-cv-00675-H-JMA (S.D. Cal. led Mar. 24, 2014), https://www.ftc.gov/system/ les/documents/cases/140409instantcheckmatecmpt.pdf. *See also* Press Release, Fed. Trade Comm'n, Two Data Brokers Settle FTC Charges at ey Sold Consumer Data without Complying with Protections Required under the Fair Credit Reporting Act (Apr. 9, 2014), https://www.ftc.gov/news-events/press-releases/2014/04/two-data-brokers-settle-ftc-charges-they-sold-consumer-data.

71 *Instant Checkmate*, No. 3:14-cv-00675-H-JMA (S.D. Cal. Apr. 1, 2014), https://www.ftc.gov/system/ les/documents/cases/140409instantcheckmateorder.pdf.

72 *See* 15 U.S.C. § 1681m(a). When using consumer reports for employment purposes, companies must also provide consumers with "pre-adverse action notices" before taking any adverse action. *See id.* § 1681b(b)(3).

73 *See*

in 2013, the FTC brought an action against Time Warner Cable because it used a consumer report to determine whether to require deposits on consumers' cable bills.[76] e complaint alleged that consumers who were charged a deposit should have received a risk-based pricing notice informing them that the charge was based on information in their consumer report. e consent order barred Time Warner Cable from future violations of the Risk-Based Pricing Rule and required the company to pay $1.9 million in civil penalties.[77] In addition, in 2015, the Commission brought an action against Sprint alleging that the

the Americans with Disabilities Act,[90] the Age Discrimination in Employment Act ("ADEA"),[91] the Fair Housing Act ("FHA"),[92] and the Genetic Information Nondiscrimination Act ("GINA").[93] These laws prohibit discrimination based on protected characteristics such as race, color, sex or gender, religion, age, disability status, national origin, marital status, and genetic information.[94]

Companies should review these laws and take steps to ensure their use of big data analytics complies with the discrimination prohibitions that may apply. This section discusses some examples of relevant considerations under these laws related to employment and credit, as highlighted in the workshop.

To prove a violation of federal equal credit or employment opportunity laws, plaintiffs typically must show "disparate treatment" or "disparate impact."[95] Disparate treatment occurs when an entity, such as a creditor or employer, treats an applicant differently based on a protected characteristic such as race or national origin.[96] Systemic disparate treatment occurs when an entity engages in a pattern or practice of differential treatment on a prohibited basis.[97] In some cases, the unlawful differential treatment could be based on big data analytics.[98] For example, an employer may not disfavor a particular protected group because big data analytics show that members of this protected group are more likely to quit their jobs within a five-year period.[99] Similarly, a lender cannot refuse to lend to single persons or offer less favorable terms to them than married persons even if big data analytics show that single persons are less likely to repay loans than married persons. Evidence of such violations could include direct evidence of the reasons for the company's choices, or circumstantial evidence, such as significant statistical disparities in outcomes for protected groups that are unexplained by neutral factors.

90 42 U.S.C. §§ 12101 *et seq*. (2014).

91 29 U.S.C. §§ 621 *et seq*. (2014).

92 42 U.S.C. §§ 3601 *et seq*. (2014).

93 42 U.S.C. §§ 2000 *et seq*. (2014). GINA also applies to health insurance.

94 A number of different agencies have the authority to enforce the various equal opportunity laws. The Equal Employment Opportunity Commission, for example, is responsible for enforcing Title VII of the Civil Rights Act of 1964 (along with the Department of Justice ("DOJ")), the Age Discrimination in Employment Act of 1967, and GINA. The Department of Housing and Urban Development and the DOJ enforce the FHA. The FTC, DOJ, and the Consumer Financial Protection Bureau ("CFPB"), among other agencies, enforce ECOA and its implementing Regulation B.

95 *See, e.g.*, Big Data Tr. 168–170 (Carol Miasko). Disparate impact claims are not permitted under Title II of GINA.

Practices that have a "disparate impact" on protected classes may also violate equal credit or employment opportunity laws.[100] While speci c disparate impact standards vary depending on the applicable law, in general, disparate impact occurs when a company employs facially neutral policies or practices that have a disproportionate adverse e ect or impact on a protected class,[101] unless those practices or policies further a legitimate business need[102] that cannot reasonably be achieved by means that have less disparate an impact.[103]

Disparate impact analysis has important implications for big data.[104] Under such an analysis, a company that avoids, for example, expressly screening job applicants based on gender and instead uses big data analytics to screen job applicants in a way that has a disparate impact on women may still be subject to certain equal employment opportunity laws, if the screening does not serve a legitimate business need

e FTC's enforcement actions include dozens of consent orders resolving alleged violations of ECOA. Some of these cases have been based on a disparate treatment theory. For example, ECOA prohibits discrimination against applicants who are receiving public assistance.[108] e Commission has brought cases against lenders that allegedly excluded public assistance income in deciding whether to extend credit.[109] Likewise, ECOA prohibits discounting or refusing to consider income on the basis of marital status.[110] e FTC has brought cases against lenders that allegedly failed to aggregate the income of unmarried joint applicants, while combining incomes for applicants who were married.[111]

e FTC also has alleged discrimination under a disparate impact legal standard under ECOA. For example, the FTC settled two cases alleging that lenders failed to appropriately monitor loan o cers whose mortgage loans resulted in minority applicants' being charged higher prices than non-Latino white applicants.[112] e Commission alleged that the statistically signi cant pricing disparities could not be

impact in subsequent lending.[116] Nevertheless, companies should proceed with caution in this area. In credit transactions,[117] Regulation B, which is the implementing regulation for ECOA, prohibits creditors[118]

of the data at issue, the size and complexity of the company's operations, and the cost of available security measures.[132]

Questions for Legal Compliance

In light of these existing laws, companies already using or considering engaging in big data

B. Special Policy Considerations Raised by Big Data Research

Workshop and seminar panelists, academics, and others have also engaged in important research in the eld of big data.[136] Some of this research has focused on how big data analytics could negatively a ect low-income and underserved populations.[137]

On the other hand, several stakeholders argue that these concerns are overstated.[141] Some emphasize that, to the extent the various steps in data mining lead to disparate impact, these issues are not new—they are inherent in any statistical analysis.[142] Other writers note that, rather than disadvantaging minorities in the hiring process, big data can help to create "a labor market that's fairer to people at every stage of their careers."[143] For example, companies can use big data algorithms to nd employees from within underrepresented segments of the population.[144] ey can also use big data to identify biases so that they can choose candidates based on merit rather than using mechanisms that depend on the reviewers' biases.[145] Furthermore, as other stakeholders have noted, big data can help "reduce the rate of 'false positive' cases that potentially make disparate treatment a problem"[146] and can help identify whether correlations exist between prices and variables such as race, gender or]TJuETk0Ɔ

continued relying on the biased data, it might have skewed road services to higher income neighborhoods. e team addressed this problem by issuing its application to city workers who service the whole city and supplementing the data with that from the public.[153] is example demonstrates why it is important to consider the digital divide and other issues of underrepresentation and overrepresentation in data inputs before launching a product or service in order to avoid skewed and potentially unfair rami cations.

2. Does your data model account for biases?

While large data sets can give insight into previously intractable challenges, hidden biases at both the collection and analytics stages of big data's life cycle could lead to disparate impact.[154] Researchers have noted that big data analytics "can reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply re ect the widespread biases that persist in society."[155] For example, if an employer uses big data analytics to synthesize information gathered on successful existing employees to de ne a "good employee candidate," the employer could risk incorporating previous discrimination in employment decisions into new employment decisions.[156] Even prior to the widespread use of big data, there is some evidence of the use of data leading to the reproduction of existing biases. For example, one researcher has noted that a hospital developed a computer model to help identify "good medical schooluEMC u/Span A

Companies should therefore think carefully about how the data sets and the algorithms they use have been generated. Indeed, if they identify potential biases in the creation of these data sets or the algorithms, companies should develop strategies to overcome them. As noted above, Google changed its interview and hiring process to ask more behavioral questions and to focus less on academic grades after discovering that replicating its existing de nitions of a "good employee" was resulting in a homogeneous tech workforce.[160] More broadly, companies are starting to recognize that if their big data algorithms only consider applicants from "top tier" colleges to help them make hiring decisions, they may be incorporating previous biases in

If you have no idea what is behind a correlation, you have no idea what might cause that correlation to break down."[166]

As another example, workshop participants discussed the fact that lenders can improve access to credit by using non-traditional indicators, e.g., rental or utility bill payment history.[167] Consumers, however, have the right to withhold rent if their landlord does not provide heat or basic sanitation services. In these instances, simply compiling rental payment history would not necessarily demonstrate whether the person is a good credit risk.[168]

the transaction or be able to correct the information used to ag them as "risky" even if the underlying information was inaccurate.[170]

In using big data to make decisions that a ect consumers' ability to complete transactions, companies should consider the potential bene ts and harms, especially where their policies could negatively a ect certain populations.

4. Does your reliance on big data raise ethical or fairness concerns?

Companies should consider performing their own assessment of the factors that go into an analytics model and balancing the predictive value of the model with fairness considerations.[171] Indeed, overreliance on the predictions of big data analytics could potentially result in a company not thinking critically about the value, fairness, and other implications of their uses of big data.[172] For example, one company determined that employees who live closer to their jobs stay at these jobs longer than those who live farther away.[173] However, another company decided to exclude this factor from its hiring algorithm because of concerns about racial discrimination, particularly since di erent neighborhoods can have di erent racial compositions.[174]

Many companies are not only considering ethical concerns with using big data, but are actively using big data to advance the interests of minorities and ght discrimination. For example, there are now recruiting tools available that match companies in search of employees with candidates who hold the necessary quali cations, but also ensure that those candidates are not limited to particular gender, racial, and experiential backgrounds.[175] Individual companies are also changing their hiring techniques to promote

170 *See* Data Brokers Report, *supra* note 7, at 53–54.

171 *See*, *e.g.*, Big Data Tr. 238–40 (TJuETnet, a.3(4(itzthaald))36uETuEMC u/Span ≠MCID 3316 ≥uEMC u/Span ≠MCID 3290 10 6 ,4a) Ba

diversity.[176] Xerox now uses an online evaluation tool developed by a data analytics rm to assess applicants, in addition to conducting interviews, to determine which applicants are most quali ed for available jobs.

V. Conclusion

Big data will continue to grow in importance, and it is undoubtedly improving the lives of underserved communities in areas such as education, health, local and state services, and employment. Our collective

Appendix:
Separate Statement of Commissioner Maureen K. Ohlhausen

Big Data: A Tool for Inclusion or Exclusion?

January 6, 2016

I support today's report on big data as a useful contribution to the ongoing policy discussion about the e ect of big data analysis on low-income, disadvantaged, and vulnerable consumers. One part of the report summarizes the concerns of several privacy advocates and academics over the potential inaccuracies of big data analytics. I write separately to emphasize the importance of evaluating these opinions in the context of market and competitive forces that a ect all companies using big data analytics.

e report details the use of big data as it a ects low-income, disadvantaged, or vulnerable consumers. Importantly, the report describes some of the many ways companies are already using big data to bene t such consumers—and others. e report also recognizes big data's massive potential bene ts. In addition, the report sketches the legal landscape implicated by big data and o ers questions that companies may nd useful as they apply big data techniques to solve their business challenges.

e report also describes certain concerns about big data tools raised by some consumer advocates and researchers. Speci cally, some fear that big data analysis will produce inaccurate or incomplete results, and that actions based on such awed analysis will harm low-income, disadvantaged, or vulnerable consumers.[1] For example, some worry that companies may use inaccurate big data analysis to deny opportunities to otherwise eligible low-income or disadvantaged consumers, or to fail to advertise high-quality lending products to eligible low-income customers.[2]

Concerns about the e ects of inaccurate data are certainly legitimate, but policymakers must evaluate such concerns in the larger context of the market and economic forces companies face. Businesses have strong incentives to seek accurate information about consumers, whatever the tool. Indeed, businesses use big data speci cally to increase accuracy. Our competition expertise tells us that if one company draws incorrect conclusions and misses opportunities, competitors with better analysis will strive to ll the gap.[3]

---

1 Fed. Trade Comm'n, Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues 8–11, 25–27 (2016). e report also references other concerns that big data analysis will be *too* accurate: companies will understand their consumers too well and misuse that data to the consumer's detriment. Market forces also constrain many such potential harms, but other such harms could actually undermine market forces. For example, the report describes concerns that unscrupulous businesses will use big data techniques to develop "sucker lists" of consumers particularly vulnerable to scams and misleading o ers. e report does a good job laying out the existing legal framework that applies to such harmful uses.

2 *Id.* at 9–11.

3 A real world example of the competitive advantages of novel but accurate application of data analytics was famously chronicled in the book (and movie) *Moneyball*. *See* Michael Lewis, Moneyball: The Art of Winning an Unfair Game (2004). Oakland's strategy succeeded precisely because it "liberated" baseball players from "unthinking prejudice rooted in baseball's traditions . . . allowing them to demonstrate their true worth." *Id.* at iiv. Each baseball franchise continually faces

Federal Trade Commission
January 2016