NOTE

DERIVED INFORMATION PROVIDED BY DATA BROKERS FOR MARKETING PURPOSES: AN ELABORATE CONSUMER PROFILE

*Yuhui Lin**

Data brokers collect personal information about consumers from a variety of sources, and they sell both the factual data and the "derived information" to businesses and individuals. "Derived information" refers to inferences about consumers made by data brokers using factual data and predictive algorithms. This information, packaged into elaborate consumer profiles, is extremely valuable to businesses. However, derived information also raises several privacy concerns. First, there is a lack of transparency in the industry. Second, consumers may face unavoidable harm due to inaccurate information. Third, derived information may facilitate discrimination based on race or gender. Unfortunately, current regulations may be inadequate to deal with these concerns.

This Note considers whether additional regulations are necessary to address issues raised by the derived information provided by data brokers for marketing purposes, and this Note recommends reliance on statutory limitations and law enforcement agencies to protect consumers' privacy interests. Part I explains how the data broker business works, and the general privacy concerns and value created by data brokers. Part II discusses the unique nature of derived information and any additional problems it may cause. Part III compares how existing laws tackle these problems and evaluates their effectiveness. Finally, Part IV recommends detailed statutory limitations. This Note recommends: (1) shifting from human-focused to behavior-focused models; and (2) limiting input variables to behavior data, while giving data brokers more leeway to process and analyze consumers' data to counteract the negative effects of the strict limitations.

* J.D., Cornell Law School, 2025. I would like to thank all editors of the *Cornell Journal of Law and Public Policy* for their work on this Note.

INTRODUCTION

Data brokers collect personal information about consumers from a variety of sources and sell both the factual data and the "derived information" to businesses and individuals.[1] "Derived information" refers to inferences about consumers made by data brokers using factual data and predictive algorithms.[2] Data brokers can even rely on completely innocuous data to infer consumers' sensitive characteristics.[3] From a marketing perspective,

---

[1] *See* FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY 3 (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [https://perma.cc/R88V-AFCL].

[2] *See id.* at ii, 19.

[3] *See, e.g.*, Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAGAZINE (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [https://perma.cc/KMF9-C4KQ] (discussing how businesses can use purchasing data alone to predict

data brokers provide significant value to their business clients by offering an elaborate consumer profile, which enables businesses to target consumers more accurately and generate more customized campaign messages based on consumer preferences.[4]

However, the nature of the derived information may cause various privacy concerns. First, regardless of the reality that data brokers are unwilling to disclose their predictive models, consumers, usually without knowledge of data analytics, may have difficulty understanding the models and how their data is used, leading to a lack of transparency in the data broker industry.[5] Second, consumers may suffer harm caused by inaccurate derived information; inaccuracies are unavoidable due to the nature of predictive algorithms.[6] In addition, consumers cannot effectively detect and correct the inaccuracies because of the lack of transparency.[7] Third, derived information may facilitate discrimination by providing previously unavailable sensitive data and seemingly innocuous inferences produced by algorithms that consider suspect features, such as race and gender.[8] Businesses may treat consumers differently based on such information.[9]

Existing regulations focus primarily on issues caused by factual data and are inadequate to deal with the concerns raised by unique features of derived information.[10] Regulating data sources is ineffective because data brokers can infer sensitive information from nonsensitive factual data. Giving consumers more control over their data does not provide them with enough protection against harm caused by derived information due to the complexity of predictive models used to produce derived information. However, allowing consumers to completely opt out of the usage of their data is also not a desirable solution. While this method may eliminate all potential harm to consumers who choose to opt out, it will damage advertisers and other consumers because of inaccuracies in the derived information. When a significant number of consumers prohibit

---

that a customer is pregnant); Paul Boutin, *The Secretive World of Selling Data About You*, NEWSWEEK (May 30, 2016), https://www.newsweek.com/secretive-world-selling-data-about-you-464789 [https://perma.cc/87WB-9F2M] (discussing how a business used online shopping data to predict consumers' health risk).

[4] FED. TRADE COMM'N, *supra* note 1, at 31.

[5] *See* Amy J. Schmitz, *Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots*,*"* 2014 MICH. ST. L. REV. 1411, 1414–15 (2014); FED. TRADE COMM'N, *supra* note 1, at 42.

[6] *See* David C. Vladeck, *Consumer Protection in an Era of Big Data Analytics*, 42 OHIO N.U. L. REV. 493, 494–95 (2016).

[7] *See id.* at 512.

[8] *See* Rebecca Lipman, *Online Privacy and the Invisible Market for Our Data*, 120 PA. ST. L. REV. 777, 782 (2016); Schmitz, *supra* note 5, at 1416–17.

[9] *See* Ashley Kuempel, Comment, *The Invisible Middleman: A Critique and Call for Reform of the Data Broker Industry*, 36 NW. J. INT'L L. & BUS. 207, 211 (2016).

[10] *See* Daniel J. Solove, *Data is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data*, 118 NW. U. L. REV. 1081 (2024).

data brokers from using their data to build predictive models and produce derived information, the aggregated input data used to train the models will likely be skewed, making the predictions more inaccurate.

This Note considers whether additional regulations are necessary to effectively address issues raised by derived information provided by data brokers for marketing purposes. This Note recommends reliance on statutory limitations and law enforcement agencies to protect consumers' privacy interests. Part I explains how the data broker business works along with the general privacy concerns and value created by data brokers. Part II discuses the unique nature of derived information and additional problems it may cause. Part III compares how existing laws tackle these problems and evaluates their effectiveness. Part IV recommends detailed statutory limitations, which focus on shifting from human-focused to behavior-focused models and limiting input variables to behavior data, while in the meantime, giving data brokers more leeway to process and analyze consumers' data to counteract the negative effects of the strict limitations.

I. Data Brokers

Data brokers collect consumers' personal information from multiple sources and then aggregate, analyze, and share the data with businesses and individual clients.[11] For instance, a data broker may collect a consumer's data from the consumer's social website and offline purchase history. The data broker can then analyze the consumer's preferences in a given category of products, so, when the consumer registers on an online shopping platform, the data broker can inform that platform which types of products likely interest the consumer. Collecting and selling consumer data is not a new business, but the increasing volume and quality of data available for analysis alongside technological advances facilitating this process continue to reshape the business today.[12]

A. How Data Broker Business Works

Data brokers collect two types of data: (1) demographic, which describes characteristics, such as gender, age, marital status, education level, and political affiliation; and (2) behavior data, which records conduct, such as purchase information and social media activities.[13] Data brokers obtain data from government sources, other publicly available sources, and commercial sources.[14] Government sources provide information on

---

[11] Fed. Trade Comm'n, *supra* note 1, at 3.

[12] *See* Staff of S. Comm. on Com., Sci. & Transp., 113th Cong., A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes 1–2 (2013) [hereinafter A Review of the Data Broker Industry].

[13] *See* A Review of the Data Broker Industry, *supra* note 11, at 13–14.

[14] Fed. Trade Comm'n, *supra* note 1, at 11.

both an individual level, such as professional licenses and real property records, and an aggregate level, such as the demographics of a particular city block.[15] Other publicly available data includes information that individuals post on the Internet and data contained in directories and press reports.[16] Commercial sources include business entities, such as retailers, magazine publishers, and financial service companies, that share their customer data.[17] Data brokers also buy information from each other.[18]

After collecting data from multiple sources, data brokers integrate the data into a single database, analyze the data, and then build predictive models to produce derived information.[19] They predict consumers' behaviors and separate consumers into segments based on their factual and predicted characteristics.[20] Their goal is to achieve a holistic view of each customer which their clients can use to customize services for their customers.[21]

Data brokers provide products in three categories: (1) risk mitigation, which helps businesses confirm consumers' identities and detect fraud;[22] (2) people search, which provides individual clients with the publicly available personal information of other individuals;[23] and (3) marketing, which enables businesses to create tailored marketing strategies for different consumers.[24]

This Note focuses on the marketing products, which include direct marketing, online marketing, and marketing analytics.[25] Under the direct marketing category, data brokers append new information to the client's data set, such as additional contact information and purchasing habits, and they provide marketing lists consisting of consumers sharing certain characteristics designated by the client.[26] Such information contains both factual information and derived information produced by predictive models, which is discussed in greater detail below.[27] Online marketing products include registration targeting, where data brokers provide data to facilitate a more customized new user experience; collaborative targeting, where data brokers analyze the user lists from the registration website and

---

15 *Id.* at 11–12.

16 *Id.* at 13.

17 *Id.* at 1, 13–14.

18 *Id.* at 14.

19 *See* Vladeck, *supra* note 6, at 498; Schmitz, *supra* note 5, at 1427.

20 FED. TRADE COMM'N, *supra* note 1, at 19.

21 *See* Schmitz, *supra* note 5, at 1427; PAM DIXON & ROBERT GELLMAN, THE SCORING OF AMERICA: HOW SECRET CONSUMER SCORES THREATEN YOUR PRIVACY AND YOUR FUTURE 8 (2014), https://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf [https://perma.cc/HNR2-6S9D].

22 FED. TRADE COMM'N, *supra* note 1, at 32.

23 *See id.* at 34.

24 *Id.* at 23.

25 *Id.*

26 *Id.* at 24–25.

27 A REVIEW OF THE DATA BROKER INDUSTRY, *supra* note 12, at 22.

the customer lists from advertisers to decide whether to advertise on the registration website; and onboarding, where data brokers add offline data into a cookie.[28] Marketing analytics products provide clients with useful insight regarding marketing strategies, such as the type of media channel to use and where advertisements should be shown.[29]

B. Privacy Concerns Raised by Data Broker Business

There are two categories of privacy harm—objective harm and subjective harm.[30] Data broker businesses may give rise to both types of privacy harm. Objective privacy harms involve the forced or unanticipated use of consumers' information against themselves, such as the government's use of sensitive personal information to limit a citizen's access to certain services.[31] Subjective privacy harm refers to the perception of unwanted observation by others, including observation of one's demographic features, behaviors, preferences, whereabouts, and inferences made based on known information.[32]

First, the data broker business lacks transparency. Generally, consumers do not know what data has been collected or how such data will be used.[33] While privacy policies may provide consumers with notice, virtually nobody reads them.[34] Besides, consumers have limited means to know whether the information held by data brokers is accurate, especially when inferences made by algorithms are involved.[35] Sometimes consumers may not even know that a marketing product with erroneous information was used against them.[36] For example, a data broker may incorrectly predict a consumer's value to a type of business based on inaccurate purchase data, and thus the consumer may obtain less favorable deals from that type of business due to their "low value." In this situation, the consumer has no way of knowing exactly why they get fewer discounts than other consumers or why they are considered a "low value" consumer because many different factors can lead to this result. The lack of transparency may aggravate subjective harm because a consumer might perceive data brokers as knowing more information about the consumer than they actually do.

---

28 FED. TRADE COMM'N, *supra* note 1, at 26–27.

29 *Id.* at 31.

30 M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131, 1142 (2011).

31 *See id.* at 1143 (providing an example that the government can leverage data mining of sensitive personal information to block a citizen from air travel).

32 *See id.* at 1144.

33 *See* A REVIEW OF THE DATA BROKER INDUSTRY, *supra* note 12, at 5.

34 Lipman, *supra* note 8, at 786.

35 A REVIEW OF THE DATA BROKER INDUSTRY, *supra* note 12, at 5; *see* Lipman, *supra* note 8, at 778.

36 *See* FED. TRADE COMM'N, *supra* note 1, at 48.

Second, consumers lack control over the compilation and use of data.[37] In addition, because of the lack of transparency, consumers may not know why they are denied a certain benefit and thus cannot take action to correct the errors in data brokers' databases.[38] The lack of control may cause objective harm because unanticipated, erroneous information can lead to the denial of benefits to consumers. It also gives rise to subjective privacy harms because, for example, consumers may perceive that advertisers possess their sensitive information based on the advertisements they receive, and they may not know how to stop advertisers from using such information.

Third, even accurate information held by data brokers may facilitate discrimination.[39] Businesses use consumer scores and segmentations provided by data brokers to determine what deals they provide to a certain consumer, leading to unequal access to benefits among consumers.[40] The models used by data brokers to generate the scores and segmentations may involve sensitive information, such as race, gender, zip code, and social status.[41] Thus, data brokers' products may foster discrimination by providing the most benefits to the wealthiest and most sophisticated consumers.[42] Therefore, discrimination may cause objective privacy harm because consumers may not be able to anticipate that their personal characteristics will cause them to receive less favorable deals.

Notwithstanding the extensive privacy harms raised by data broker businesses, seeking relief from the courts is difficult for consumers. To satisfy the injury-in-fact element of the standing requirement, the harm must be "concrete."[43] Courts generally consider privacy harms compensable only when "cognizable," "actual," "specific," "material," "fundamental," or "special."[44] Insubstantial inaccuracies or failure to provide required notice to a consumer in violation of statute are unlikely to cause harm or present a material risk of harm.[45] Thus, it is difficult for consumers to protect against at least the harm caused by the lack of transparency and the standing requirement. The obstacles to bringing discrimination claims is discussed in Part II of this Note.

---

37 A REVIEW OF THE DATA BROKER INDUSTRY, *supra* note 12, at 5.

38 *See* FED. TRADE COMM'N, *supra* note 1, at 48.

39 *See* Lipman, *supra* note 8, at 782; A REVIEW OF THE DATA BROKER INDUSTRY, *supra* note 12, at 6.

40 *See* Schmitz, *supra* note 5, at 1411; A REVIEW OF THE DATA BROKER INDUSTRY, *supra* note 12, at 6.

41 *See* Schmitz, *supra* note 5, at 1411.

42 *See Id.*

43 Spokeo, Inc. v. Robins, 578 U.S. 330, 342 (2016).

44 Calo, *supra* note 30, at 1132.

45 *Spokeo*, 578 U.S. at 342 (finding no concrete harm where an agency fails to provide required notice regarding the use of consumer information but where the information is accurate or the only inaccuracy is an incorrect zip code).

C. Values Created by Data Brokers' Marketing Products

Data brokers' marketing products benefit their business clients by improving marketing efficiency through precise marketing strategies and facilitating product and service improvements. An elaborate consumer profile enables businesses to target consumers more accurately for advertising purposes and better refine their campaign messages based on consumer preferences.[46] For example, advertisers can emphasize a product's quality more than its price in advertisements sent to high-end consumers. Aggregated consumer data and analytics insights may help businesses enhance their products and services according to consumer preferences. Further, the data and analytics may better equip businesses to make more general marketing decisions, such as which type of media channel to use and where advertisements should be shown.[47]

Individual consumers also benefit from data brokers' marketing products. Precise marketing strategies allow consumers to more easily find goods and services that meet their preferences and needs.[48] Besides, consumers are "compensated" for the use of their data by accessing various online platforms' services for free, as those platforms can generate substantial revenues from selling targeted advertisements.[49] While the platforms can still sell advertisements without the information provided by data brokers, the advertising campaign will be less precise, lowering the advertisers' marketing efficiency and making them less willing to pay for the advertisements, which will lower the platforms' revenue and subsequently influence the free services consumers receive.

II. DERIVED INFORMATION SOLD BY DATA BROKERS

A. Derived Information

Derived information, a type of data provided by data brokers, is the inference that data brokers make about consumers ("data subjects") based on their factual data. [50] Derived information falls into three categories: (1) consumers' interest in particular products, (2) consumers' general characteristics and interests, and (3) consumer ratings.[51]

Data brokers use algorithms to predict consumers' interests based on their characteristics and behaviors.[52] Data brokers first analyze the data of consumers who already bought a product to identify characteristics and behaviors shared among them and then predict that other consumers

---

46 FED. TRADE COMM'N, *supra* note 1, at 31.
47 *Id.*; Schmitz, *supra* note 5, at 1419.
48 FED. TRADE COMM'N, *supra* note 1, at 47.
49 *See* Lipman, *supra* note 8, at 784.
50 FED. TRADE COMM'N, *supra* note 1, at ii.
51 *See* Vladeck, *supra* note 6, at 496; Schmitz, *supra* note 5, at 1414.
52 *See* FED. TRADE COMM'N, *supra* note 1, at 19.

with similar characteristics and behaviors will be interested in that product.[53] Data brokers also create segments based on consumers' general characteristics, such as "Health & Wellness Interest," which can be used for various marketing purposes.[54] Consumer ratings assess each consumer's likely value to businesses and facilitate decision-making concerning how to treat different consumers.[55]

Some derived data may involve very sensitive information, such as "Expectant Parent," "Diabetes Interest," "Financially Challenged," and information related to their sexual orientations.[56] On the one hand, seemingly innocuous labels may come from sensitive factual data, such as "Urban Scramble," which is predicted based on ethnicity and income levels.[57] On the other hand, algorithms may enable data brokers to predict very sensitive information based on completely non-sensitive and innocuous data.[58] For example, Target predicted whether a customer is pregnant merely based on their purchase data.[59] This nature of derived information makes regulating data brokers' processing and selling of potentially sensitive data challenging. Further, the scope of "sensitive information" that needs to be strictly regulated is hard to define.

B. Problems Caused by the Sale and Use of Derived Information

Compared with factual information, the unique nature of derived information may aggravate all the existing concerns for the data broker industry. First, the lack of transparency becomes a more significant problem because, normally, data brokers provide a consumer with access to his factual data but not to the derived information, making the consumer unaware of how he is segmented.[60] Besides, the algorithms behind derived information may constitute trade secrets, which data brokers have no legal duty to disclose so long as the data is not used for certain highly regulated purposes.[61] While some data brokers may provide a consumer with ratings or some general interest categories associated with them, such as "Travel Enthusiast" or "Green Consumer," such derived information is impossible to decipher without an in-depth understanding of data analytics.[62]

---

53 *Id.*

54 *See id.* at 20.

55 Schmitz, *supra* note 5, at 1414, 1428.

56 *See* FED. TRADE COMM'N, *supra* note 1, at 47; Vladeck, *supra* note 6, at 500.

57 *See* FED. TRADE COMM'N, *supra* note 1, at 47; Schmitz, *supra* note 5, at 1413.

58 *See* Duhigg, *supra* note 3.

59 *Id.*

60 FED. TRADE COMM'N, *supra* note 1, at 42.

61 *See* Schmitz, *supra* note 5, at 1415 (suggesting that data brokers may have a duty to disclose if their data are used for determining credit, insurance, or employment).

62 *See id.*, at 1413.

Second, derived information is less accurate than factual data, so discovering and correcting inaccuracies in derived information is more difficult. The foundation of the predictive models, or the factual information, is inaccurate, as one of the largest data brokers admitted that thirty percent of a data subject's profile may be wrong.[63] Also, no algorithm can be completely free of error because algorithms are based on correlations rather than hard facts, which means that unfairly characterized consumers are unavoidable collateral damage even though the algorithm is reasonably accurate.[64] For example, if the algorithm finds that eighty percent of consumers with characteristic X bought product Y, it will predict that anyone with characteristic X will buy product Y. Therefore, twenty percent of consumers would be falsely categorized. Thus, the inaccuracies in the factual data are amplified in the derived data. Notwithstanding the growing inaccuracies, the lack of transparency makes consumers even less likely to detect errors, making it even harder to achieve accuracy.[65]

Last, the sale and application of derived information may further facilitate discrimination. Derived information involves vast amounts of personal information that was not previously available. Businesses can use potentially incorrect data to make discriminatory decisions.[66] Consumers may be unaware of the discrimination because of the lack of transparency and, thus, have no means to seek relief from the courts.[67] Besides, some derived information, especially consumer ratings, may obscure and even justify discrimination.[68] For example, businesses subject their consumers to unequal access to information, differential pricing, and predatory practices based on the consumers' ratings or scores provided by data brokers.[69] Providing disparate treatment according to each consumer's potential value to a business may constitute a legitimate rationale. However, it may also lead to a disproportionately adverse effect on certain groups because the algorithms used to calculate consumer ratings may factor in race, gender, and other suspect considerations that may foster discrimination.[70] Nevertheless, since the ratings are generated in the black box of the algorithms, proving discriminatory intent or impact, which is required in discrimination cases, is very difficult.[71] As a result, the derived information provided by data brokers may in effect impose disparate impacts on

---

63 *Id.*, at 1428; Lipman, *supra* note 8, at 782.

64 *See* Vladeck, *supra* note 6, at 494.

65 *See* Vladeck, *supra* note 6, at 512.

66 *See* Lipman, *supra* note 8, at 782.

67 *See id.*, at 782.

68 Kuempel, *supra* note 9, at 210.

69 *Id.*, at 211.

70 *See* Vladeck, *supra* note 6, at 513; Schmitz, *supra* note 5, at 1417.

71 Vladeck, *supra* note 6, at 514 (arguing that discriminatory intent or impact can hardly be detected because algorithms process data in a fluid manner and the factors that are deemed significant can change over time).

consumers based on factors that may otherwise give rise to discrimination claims. Furthermore, the very nature of predictive algorithms may render their application discriminatory. As discussed above, algorithmic decision-making is based on correlations.[72] The decision generated by mere correlations may be a stereotype because it assumes that people sharing certain characteristics must behave in a particular way. Where the classification is merely based on gender, it may constitute discrimination if it rests on impermissible stereotypes, even when some statistics support the generalization.[73] A classification should not be deemed compliant with discrimination law merely because it factors in suspect characteristics together with innocuous elements. So derived information produced by predictive algorithms with input variables concerning race, gender, or other sensitive characteristics may also constitute discrimination. In addition, this issue is further complicated because data brokers may still predict consumers' race and gender using non-sensitive data and then put the predicted data into the algorithm even if the use of race and gender to train algorithms is forbidden.

III. LAWS TO REGULATE DERIVED INFORMATION AND THEIR INADEQUACY

In the United States, laws specifically directing to data brokers mainly focus on the registration requirement rather than the data used by data brokers.[74] Despite those laws, to regulate data used for marketing purposes, several federal laws impose limits on data sources.[75] Unfortunately, they provide limited relief to the privacy concerns raised by derived information because data brokers can use algorithms to predict information similar to that provided by the regulated sources.[76] Thirteen states enacted data privacy laws as of November 2023,[77] and more

---

72 *Id.*, at 513.

73 *See, e.g.*, J.E.B. v. Alabama *ex rel.* T.B., 511 U.S. 127, 139 n.11 (1994).

74 *See* CAL. CIV. CODE §§ 1798.99.80–1798.99.89 (2019).

75 *See, e.g.*, 15 U.S.C. § 1681b (regulating consumer reports); 45 C.F.R. § 164.512 *et seq.* (regulating individually identifiable health information); 15 U.S.C. § 6802 (regulating how financial institutions collect and disclose nonpublic personal information).

76 *See* Solove, *supra* note 10, at 1081.

77 *See* California Consumer Privacy Act of 2018, CAL. CIV. CODE §§ 1798.100 to 1798.99.100 (2018); Colorado Privacy Act, COLO. REV. STAT. ANN. §§ 6-1-1301 to 6-1-1314 (2021); Connecticut Data Privacy Act, CONN. GEN. STAT. ANN. §§ 42-515 to 42-527 (2022); Delaware Personal Data Privacy Act, DEL. CODE ANN. tit. 6, §§ 12D-101 to 12D-111 (2023) (effective Jan. 1, 2025); Florida Digital Bill of Rights Act, FLA. STAT. §§ 501.701–501.722 (2023); Indiana Consumer Data Protection Act, IND. CODE § 24-15 (2023) (effective Jan. 1, 2026); Iowa Consumer Data Protection Act, IOWA CODE ANN. § 715D. (2023) (effective Jan. 1, 2025); Montana Consumer Data Privacy Act, MONT. CODE ANN. §§ 30-14-2801 to 30-14-2817 (2023); Oregon Consumer Privacy Act, OR. REV. STAT. ANN. § 646A.570–646A.592 (2023); Tennessee Information Protection Act, TENN. CODE ANN. § 47-18-3301 to 47-18-3315 (2023) (effective July 1, 2025); Texas Data Privacy and Security Act, TEX. BUS. & COM. CODE ANN.

states have followed suit.[78] These state laws mainly focus on enhancing transparency by requiring notice to consumers, giving consumers more control over their data by allowing them to delete and correct their data and opt out of the data processing and sharing program, and addressing privacy concerns by limiting the scope of the sensitive data that can be analyzed and shared.[79] While these regulations may be effective in regulating factual data, they are inadequate to deal with the issues raised by derived information.

A. Regulating Data Sources

The federal government has enacted laws limiting the disclosure of data by certain sources that have sensitive information, such as medical conditions and financial status.[80] For example, the Health Insurance Portability and Accountability Act prohibits healthcare providers from disclosing individually identifiable health information, and the Gramm-Leach-Bliley Act prohibits financial institutions from sharing consumers' personal financial information with an unaffiliated third party without notice.[81] As a result, data brokers cannot obtain certain sensitive information from those sources, thus ameliorating consumers' privacy concerns and reducing the risk of discrimination based on some protected characteristics.[82]

However, the application of derived information may make this solution ineffective. For instance, data brokers can use algorithms to infer similar sensitive information as that provided by regulated sources

---

§§ 541.001–541.205 (2023); Utah Consumer Privacy Act, UTAH CODE ANN. § 13-61-101 to 13-61-404 (2022); Virginia Consumer Data Protection Act, VA. CODE ANN. § 59.1-575 to 59.1-585 (2021) (effective Jan. 1, 2025).

78 While this Note underwent final stages of publication, more states enacted data privacy laws. *See* F. Paul Pittman, Hope Anderson & Abdul M. Hafiz, *US Data Privacy Guide,* WHITE & CASE (July 2, 2024) https://www.whitecase.com/insight-our-thinking/us-data-privacy-guide?s=data%20privacy%20guide [https://.cc/S56U-XVBV] (providing a state-by-state analysis of data privacy laws). The article explains that Kentucky, Maryland, Minnesota, Nebraska, New Hampshire, New Jersey, and Rhode Island have enacted similar data privacy laws. *Id.*

79 *See* statutes cited *supra* note 77.

80 Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936, 1992; Gramm-Leach-Bliley Act, 15 U.S.C. § 6801 (1999).

81 *See* Health Information Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat.1936; 45 C.F.R. § 164.512 *et seq.*; Gramm-Bliley Act, 15 U.S.C. § 6801 (1999).

82 *See generally* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 16, 47 (2012) ("[W]hen health or children's information is involved, for example, the likelihood that data misuse could lead to embarrassment, discrimination, or other harms is increased."); JUSTIN SHERMAN, DUKE UNIV., SANFORD CYBER POL'Y PROGRAM, DATA BROKERS AND SENSITIVE DATA ON U.S. INDIVIDUALS 9–10 (2021) ("Data brokers . . . hold highly sensitive data on U.S. individuals such as race, ethnicity, gender, sexual orientation, immigration status, income level and political preferences and beliefs . . . that can be used to directly undermine individual's civil rights.").

from innocuous purchasing or online browsing behavior data.[83] The only difference is that the "predicted features" are less accurate than those held by the strictly regulated sources.[84] The growing inaccuracy in the data may increase the likelihood of discrimination and cause more objective harm to consumers because more consumers may be falsely deprived of the benefits or interests that they deserve. Furthermore, defining which data source is deemed to "have sensitive consumer information" is difficult because many sources may not have literally sensitive information, but they may have data that implies sensitive characteristics, such as purchasing history that involves food and dietary supplements particularly suitable for people with certain diseases. Therefore, the predictive power of algorithms and the vast amount of innocuous data make regulating data sources inadequate to deal with the privacy and discrimination concerns raised by derived information in the modern data broker industry.

B. Providing Consumers with Notice and Access to Their Own Data

Requiring data brokers to provide consumers with notice and access to their data helps enhance transparency.[85] All thirteen states with data privacy laws have a privacy notice requirement mandating every business collecting consumers' personal information ("data controller") to disclose at least the categories of personal information it collects, the purpose for processing the information, the categories of third parties with whom it discloses personal information, and the categories of personal information it discloses to third parties.[86] Additionally, Florida requires the disclosure of specific pieces of personal information, rather than mere categories, that a data controller has collected about consumers or shared with third parties.[87] Also, Florida and Delaware require the disclosure of a list of specific third parties to whom personal data is shared.[88] Furthermore, if a data controller sells personal data to third parties or processes the data for targeted marketing aims, eleven states explicitly require consumer

---

[83] *See* FED. TRADE COMM'N, *supra* note 1, at 31.

[84] *See* RAHUL KANWAL & KEVIN WALBY, UNIV. OF WINNIPEG CTR. FOR ACCESS TO INFO. & JUST., TRACKING THE SURVEILLANCE AND INFORMATION PRACTICES OF DATA BROKERS: A REPORT 11 (2024).

[85] *See* FED. TRADE COMM'N, *supra* note 82, at 60.

[86] *See* CAL. CIV. CODE §§ 1798.110(c), 1798.115(a) (2024); COLO. REV. STAT. ANN. § 6-1-1308(1)(a) (2024); CONN. GEN. STAT. § 42-520(c) (2024); DEL. CODE ANN. tit. 6, § 12D-106(c) (2024) (effective Jan. 1, 2025); FLA. STAT. § 501.711(1) (2023); IND. CODE § 24-15-4-3 (2024) (effective Jan. 1, 2026); IOWA CODE ANN. § 715D.4(5) (2024) (effective Jan. 1, 2025); MONT. CODE ANN. § 30-14-2812(5) (2023); OR. REV. STAT. ANN. § 646A.578(4) (2024); TENN. CODE ANN. § 47-18-3305(c) (2024) (effective July 1, 2025); TEX. BUS. & COM. CODE ANN. § 541.102(a) (2023); UTAH CODE ANN. § 13-61-302(1)(a) (2024); VA. CODE ANN. § 59.1-578(C) (2024) (effective Jan. 1, 2025).

[87] FLA. STAT. § 501.711(1) (2023).

[88] *See id.*; DEL. CODE ANN. tit. 6, § 12D-104(a)(5) (2024) (effective Jan. 1, 2025).

notification.[89] Among the eleven states, Oregon is the only state that requires data controllers to describe the processing of personal data for purposes of targeted advertising.[90]

Regarding the right to access, all thirteen states give data subjects the right to obtain a copy of their personal data in the possession of data controllers, though variations exist in the scope of the accessible data.[91] Seven states limit the data subjects' access merely to the personal data that has been provided to the controller by the data subjects themselves.[92] Since data brokers do not collect data directly from consumers, their data may fall out of "the data provided by data subjects," thus making this statute inapplicable to data brokers.[93] Even if this statute governs the data possessed by data brokers, derived information may still fall outside the scope of this requirement, as it consists of predictions made by data brokers based on data from various sources, including public sources completely independent from consumers' control.[94] The remaining six states allow a data subject to request a copy of all his or her personal data possessed by the controller, but they explicitly provide that the statute does not require the controller to reveal any "trade secret."[95] The models and algorithms used to produce the derived information may constitute a "trade secret,"[96]

---

89 VA. CODE ANN. § 59.1-578(D) (2025); COLO. REV. STAT. § 6-1-1303(1)(b) (2024); UTAH CODE ANN. § 13-61-302(1)(b) (2023); CONN. GEN. STAT. § 42-520(d) (2023); IOWA CODE § 715D.4(6) (2023); IND. CODE ANN. § 24-15-4-4 (2023); TENN. CODE ANN. § 47-18-3204(d) (2023); MONT. CODE ANN. § 30-14-2812(5) (2023); TEX. BUS. & COM. CODE § 541.103 (2024); OR. REV. STAT. ANN. § 646A.578(4) (2023); DEL. CODE ANN. tit. 6, § 12D-106 (2025) (effective Jan. 1, 2025).

90 OR. REV. STAT. ANN. § 646A.578(4)(h) (2023) ("[A] controller shall provide to consumers a reasonably accessible, clear and meaningful privacy notice that . . . [p]rovides a clear and conspicuous description of any processing of personal data in which the controller engages for purposes of targeted advertising.").

91 *See* CAL. CIV. CODE §§ 1798.110(c), 1798.115(a) (2024); VA. CODE ANN. § 59.1-578(C) (2025); COLO. REV. STAT. § 6-1-1308(1)(a) (2024); UTAH CODE ANN. § 13-61-302(1)(a) (2023); CONN. GEN. STAT. § 42-518(a)(4) (2023); IOWA CODE § 715D.4(5) (2024); IND. CODE. § 24-15-4-3 (2023); TENN. CODE ANN. § 47-18-541.102(d); MONT. CODE. ANN. 30-14-2812(5) (2023); TEX. BUS. & COM. CODE § 541.102 (2024); OR. REV. STAT. ANN. § 646A.578(4) (2023); DEL. CODE. § 12D-106(d) (2025).

92 *See* CAL. CIV. CODE § 1798.130(a)(3)(B)(iii) (2018); VA. CODE ANN. § 59.1-578(A)(4) (2021); UTAH CODE ANN. § 13-61-201(3) (2022); IOWA CODE § 715D.3(1)(c) (2023); IND. CODE § 24-15-3-1(b)(4) (2023); TENN. CODE ANN. § 47-18-3203(a)(2)(D) (2023); TEX. BUS. & COM. CODE § 541.051(b)(4) (2023).

93 Justin Sherman, *People Search Data Brokers, Stalking, and 'Publicly Available Information' Carve-Outs*, LAWFARE (Oct. 30, 2023), https://www.lawfaremedia.org/article/people-search-data-brokers-stalking-and-publicly-available-information-carve-outs [https://perma.cc/SN3U-URP4].

94 *See* Kelly Drye, *Mounting Focus on Data Brokers: Is More Regulation Coming?*, KELLEY DRYE (Aug. 24, 2023), https://www.kelleydrye.com/viewpoints/blogs/ad-law-access/mounting-focus-on-data-brokers-is-more-regulation-coming [https://perma.cc/AM76-SCFL].

95 *See* COLO. REV. STAT. § 6-1-1306(1)(e) (2021); CONN. GEN. STAT. ANN. § 42-518(a) (2022); MONT. CODE ANN. § 30-14-2808(1)(d) (2023); FLA. STAT. § 501.716(4) (2023); OR. REV. STAT. ANN. § 646A.574(3) (2023); DEL. CODE ANN. tit 6, § 12D-104(a)(4) (2023).

96 *See* Schmitz, *supra* note 5, at 1447.

so it is unclear whether disclosing the derived information, the product of the trade secret, could reveal the underlying models and thus be exempt from the disclosure requirement.

Admittedly, the privacy notice gives consumers a better opportunity to understand what information data brokers have and how they process and use the information, which can help consumers make informed choices.[97] Besides, the notice requirement can discourage data brokers from engaging in practices that are likely to raise serious privacy concerns and force them to take more responsibility for the data they possess.[98]

However, these regulations provide inadequate notice to consumers when derived information is involved. First, due to the broad scope and the complex process of producing derived information, merely disclosing categories of personal information may not give consumers sufficient notice about the amount of information possessed by data brokers and how their data is processed and used. For example, a data broker may merely notify consumers that it predicts their preferred features of food, but consumers may not be able to anticipate that the data broker also predicts their health concerns for the purpose of understanding their purchasing preferences. Similarly, while consumers are aware that their information is used in targeted advertising, they may not truly understand how their data is used in this process because of a lack of explanation. For instance, a data broker may state that it uses consumers' purchasing data in targeted advertising, but consumers may not expect that an advertiser provides the most favorable deal to consumers with a purchasing amount in a certain range because those consumers are the most likely to be influenced by deals.

Second, requiring the disclosure of derived information's production process may not be an effective solution either. Providing more detailed information may, in effect, overload consumers and increase their confusion given the complexity of predictive algorithms used to produce derived information.[99] Also, a too-detailed notice may even enlarge the discriminatory effects of using derived information because the more sophisticated consumers are more likely to gain an adequate understanding of how their data is used and thus can better protect themselves, leaving their less sophisticated counterparts, who are more likely to be victims of discrimination, largely unprotected.[100]

Third, even though consumers can hardly rely on their right to access to obtain derived information from data brokers, even if they obtain all derived data, consumers may not understand the meaning and

---

97 *See* Lipman, *supra* note 8, at 787.

98 *Id.*

99 FED. TRADE COMM'N, *supra* note 82, at 61.

100 *See* Schmitz, *supra* note 5, at 1462–63.

logic behind the data. For example, consumers may receive many scores without any benchmark or seemingly innocuous tags merely describing their characteristics. Even a consumer with knowledge of data analytics will not know exactly how those scores and tags may affect the deals they receive from advertisers.[101]

Last, seeking relief from courts pursuant to this statute may be hard because the mere lack of notice may not meet the damage requirement.[102] Therefore, current laws regarding consumers' right to notice and right to access do not offer adequate solutions to the issues raised by derived information.

C. Providing Consumers with Control of Their Information

To enhance consumers' control over the compilation and use of their personal information, the thirteen states' data privacy laws give consumers the right to correct inaccurate data and delete data in possession of controllers and impose more limits on the processing and disclosure of "sensitive information."[103] Nevertheless, the characteristics of derived information discount the performance of these laws.

Except Utah and Iowa, all states with data privacy laws confer consumers the right to correct inaccurate personal information relating to themselves, taking into account "the nature of the personal information and the purposes of the processing of the personal information."[104] Notably, Indiana limits the right to only the "personal data that the consumer previously provided."[105] As discussed above, personal information possessed by data brokers and derived information may not fall within the scope of consumer-provided data, making Indiana consumers incapable of correcting at least their erroneous derived information held by data brokers.[106]

For consumers in the ten states offering the right to correct, while correcting inaccurate data could reduce their risk of being falsely deprived

---

101 Timothy Morey, Theodore Forbath & Allison Schoop, *Customer Data: Designing for Transparency and Trust*, HARV. BUS. REV., May 2015, at 96.

102 *See* Spokeo, Inc. v. Robins, 578 U.S. 330, 342 (2016).

103 CAL. CIV. CODE § 1798.106(a); VA. CODE ANN. § 59.1-577(A)(2) (2021); COLO. REV. STAT. § 6-1-1306(1)(c) (2021);CONN. GEN. STAT. ANN. § 42-518(a)(2) (2023); IND. CODE § 24-15-3-1(b)(2) (2023); TENN. CODE ANN. § 47-18-3202 (2023); MONT. CODE ANN. § 30-14-2808(1) (2023); TEX. BUS. & COM. CODE § 541.051(b)(2) (2023); FLA. STAT. § 501.705(2)(b) (2023); OR. REV. STAT. ANN. § 646A.574(1) (2023); DEL. CODE tit. 6, § 12D-104(a) (2023); UTAH CODE ANN. § 13-61-201 (2022); IOWA CODE § 715D.3(1)(b) (2023).

104 CAL. CIV. CODE § 1798.106(a); VA. CODE ANN. § 59.1-577(A)(2) (2021); COLO. REV. STAT. § 6-1-1306(1)(c) (2021);CONN. GEN. STAT. ANN. § 42-518(a)(2) (2023); IND. CODE § 24-15-3-1(b)(2) (2023); TENN. CODE ANN. § 47-18-3202 (2023); MONT. CODE ANN. § 30-14-2808(1) (2023); TEX. BUS. & COM. CODE § 541.051(b)(2) (2023); FLA. STAT. § 501.705(2)(b) (2023); OR. REV. STAT. ANN. § 646A.574(1) (2023); DEL. CODE tit. 6, § 12D-104(a) (2023).

105 IND. CODE § 24-15-3-1(b)(2) (2023).

106 *Id.*

of benefits and interests that should have belonged to them, they can hardly deal with the inaccuracies in derived information, which are more common than those in factual data because of the unavoidable false predictions made by algorithms. First, detecting inaccuracies in derived information is hard because the data could be a numerical score or a phrase that makes no sense to consumers, such as "Urban Scramble."[107] Second, many types of derived information are about data subjects' preferences, tendencies, and possibilities of performing in a certain way, making it hard to prove falsity.[108] For example, algorithms may predict that a consumer is likely to be interested in a particular product, but the consumer's not buying that product cannot demonstrate that the prediction is wrong because the consumer may choose not to buy the product due to other reasons and may buy it in the future. Thus, even if a consumer feels that a piece of derived information is inaccurate, demonstrating the inaccuracy and requesting the data broker to correct it may be difficult.

All thirteen states confer on data subjects the right to request a data controller delete personal data about the data subject[109], while three states limit the scope to the data provided by the data subject.[110] Indeed, by deleting the personal data completely from the data brokers' databases, consumers may eliminate the risk that their data is used against themselves, thus greatly reducing objective privacy harms.

However, this solution may lead to many undesirable consequences. First, the deletion of data may lower the accuracy of predictive models and make derived information more inaccurate, which may cause more consumers to be falsely denied benefits and interests. Consumers who have exercised their right to delete their personal data may share some common characteristics, such as receiving higher education, rather than being randomly distributed.[111] After the deletion of the data, data brokers have no opportunity to analyze the characteristics of the consumers who have requested to be deleted, so when data brokers build models to predict derived information, they will have to assume that the remaining consumers are a randomly selected sample of the general public.[112] Thus, the input

---

107 *See* Schmitz, *supra* note 5, at 1470.

108 *See id.*, at 1452.

109 *See* CAL. CIV. CODE § 1798.105(a) (2018); VA. CODE ANN. § 59.1-577(A)(3) (2021); COLO. REV. STAT. § 6-1-1306(1)(d) (2021); UTAH CODE ANN. § 13-61-201(2) (2022); CONN. GEN. STAT. ANN. § 42-518(a)(3)(2022); IOWA CODE § 715D.3(1)(b) (2023); IND. CODE § 24-15-3-1(b)(3) (2023); TENN. CODE ANN. § 47-18-3203(a)(2)(C) (2023); MONT. CODE ANN. § 30-14-2808(1)(c) (2023); TEX. BUS. & COM. CODE § 541.051(b)(3) (2023); FLA. STAT. § 501.173(5)(a) (2023); OR. REV. STAT. ANN. § 646A.574(1)(c) (2023); DEL. CODE ANN. tit. 6, § 12D-104(a)(3) (2023).

110 *See* CAL. CIV. CODE § 1798.105(a) (2018); UTAH CODE ANN. § 13-61-201(2) (2022); IOWA CODE § 715D.3(1)(b) (2023).

111 *See* Schmitz, *supra* note 5, at 1467 (arguing that consumers with the least education and resources are less likely to benefit from the control rights conferred by law).

112 *See id.*

data used to train the models is likely to be skewed because nonrandomly selected consumers are removed, which will make the predictions less accurate.[113] As a result, the remaining consumers who have not deleted their data are more likely to suffer harm based on inaccurate predictions.

Second, the inaccurate derived information and the decreased amount of data available for use in targeted marketing may negatively affect advertisers and data brokers. Without enough accurate personal information about consumers, advertisements will be less-accurately targeted, lowering the marketing efficiency of advertisers.[114] Thus, advertisers will be less willing to pay for the data obtained from data brokers, which means that the data broker industry will suffer.

Third, consumers themselves will be harmed. Despite the convenience provided by customized advertising materials discussed in Part I, consumers trade their data to obtain free access to services provided by online platforms.[115] A declining data broker market will diminish the value of consumers' data, making consumers lose free access to services previously available to them.

Except for Florida, the remaining twelve states impose stricter limitations on the processing of "sensitive information" than other information. Three states require data controllers to give consumers "clear notice" and "an opportunity to opt out" before processing "sensitive information,"[116] while the other nine states require affirmative consumers' consent.[117] Nevertheless, states' definitions of "sensitive information" vary. While all twelve states agree that a consumer's racial or ethnic origin, religious beliefs, mental or physical diagnosis, sexual orientation, citizenship or immigration status, and genetic or biometric data that may be processed for the purpose of uniquely identifying a natural person are "sensitive information," they have different rules regarding whether precise geolocation, philosophical beliefs, union membership, health conditions, sex life, national origin, status as transgender or nonbinary, and status as a victim of crime constitute "sensitive information" for purposes of their data privacy laws.[118] Notably, California exempts "publicly available"

---

113 *See* Vladeck, *supra* note 6, at 494–95.

114 *See* Lipman, *supra* note 8, at 784.

115 *Id.*

116 *See* Cal. Civ. Code § 1798.121(a) (2018); Utah Code Ann. § 13-61-302(3)(a) (2022); Iowa Code § 715D.4(2) (2023) (effective Jan. 1, 2025).

117 *See* Va. Code Ann. § 59.1-578(A)(5) (2024); Colo. Rev. Stat. § 6-1-1308(7) (2023); Conn. Gen. Stat. § 42-520(a)(4) (2023); Ind. Code. § 24-15-4-1(5) (2023); Tenn. Code Ann. § 47-18-3204(a)(6) (2023); Mont. Code Ann. § 30-14-7(2)(b) (2023); Tex. Bus. & Com. Code § 541.101(b)(4) (2023); Or. Rev. Stat. § 5(2)(b) (2023); Del. Code Ann. tit. 6, § 12D-106(a)(4) (2023).

118 *See* Cal. Civ. Code § 1798.140(ae) (2024); Va. Code Ann. § 59.1 -575 (2024); Colo. Rev. Stat. § 6-1-1303(24) (2024); Utah Code Ann. § 13-61-101(32)(a) (2023); Conn. Gen. Stat. Ann. § 42-515(27) (2023); Iowa Code § 715D.1(26) (2023); Ind. Code Ann. 24-15-2-28 (2023); Tenn. Code Ann. § 47-18-3302(26) (2023); Mont. Code Ann. § 30-14-2802(28)

personal information from "sensitive information."[119] The information is deemed available to the public if the means of access are widely known, and if there are no warnings, encryptions, password requests, or other indicia of intended privacy.[120]

Under the "publicly available" exemption, consumers' profiles on social media that have been made visible to the general public may constitute "publicly available" information and fall within the exemption under California law.[121] Consumers' profiles may include information that would otherwise be "sensitive," such as racial or ethnic origin, which could in effect enable data brokers to process such information without strict limitations.

Even without considering the "publicly available" exemption, the strict limitations on the processing of "sensitive information" are inadequate to deal with the privacy concerns raised by derived information. First, while states can give notice to their citizens regarding the states' definition of "sensitive information," the variations of the definitions among the states indicate that people hold different opinions about what data is "sensitive" and causes invasions of their privacy.[122] Thus, some consumers may have a broader scope of data that they consider "sensitive" than the scope defined by the state. For example, a consumer may feel that tracking his precise location invades his privacy, even though his state does not define precise geolocation as "sensitive information." As a result, the law may not give those consumers sufficient control over the information that they believe to be "sensitive," making the law inadequate in addressing their subjective privacy harm.

Second, defining which derived information is "sensitive" is difficult. Algorithms can infer "sensitive information" from non-sensitive data.[123] Then, should the "predicted sensitive information" count as "sensitive information" within the meaning of the statute of the data privacy law? If the "predicted sensitive information" still falls within the scope of "sensitive information," what if the derived information uses mere neutral words to hide its predictions on sensitive characteristics? What

---

(2023); TEX. BUS. & COM. CODE § 541.001(29) (2023); OR. REV. STAT. § 646A.593 (2023); DEL. CODE tit. 6, § 12D-102(30) (2023).

119 CAL. CIV. CODE § 1798.140(ae)(3) (2018).

120 *See* hiQ Labs, Inc. v. LinkedIn Corp., 31 F.4th 1180, 1200 (9th Cir. 2022) (quoting H.R. Rep. No. 99-647, at 62 (1986)).

121 *See id.* at 1200–01 (finding that data on user's LinkedIn profile may constitute "publicly available data" when not demarcated as private).

122 *Compare* CAL. CIV. CODE § 1798.140(ae)(3) (2024) (exempting publicly available information from sensitive information), *with* OR. REV. STAT. § 646A.593(1)(c)(B)(iii) (2023) (exempting data generated by utility services from sensitive information).

123 *See* Daniel J. Solove, *Data is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data*, 118 Nw. U. L. Rev. 1081, 1084 (2024) ("[P]owerful machine learning algorithms facilitate inferences about sensitive data from nonsensitive data.").

if the derived information is inferred from multiple innocuous factors in combination with the "predicted sensitive information?" What if the derived information's predictive model only considers the data that was used to predict "sensitive information" rather than directly considering the "predicted sensitive information?" From a practical perspective, drawing a clear line to define "sensitive information" is extremely difficult, if not completely infeasible.

Moreover, the lack of sufficient understanding of derived information caused by the complexity of predictive algorithms, as discussed previously, may further discount the effectiveness of consumers' power to control their data because consumers will not be able to detect the inaccuracies in their derived data, realize when a complete deletion of their data is necessary, nor decide when and to what extent they should request to limit the processing of their sensitive information.

D. Providing Consumers with the Right to Opt Out

Providing consumers with the right to opt out is a milder method to enhance their control over their data rather than allowing them to completely delete their information. This method imposes limitations on data brokers' use of consumers' data for certain purposes.[124] Three of the thirteen states only allow consumers to opt out of the selling or sharing of personal data,[125] while the other ten states permit consumers to opt out of any processing of personal data for certain purposes, such as targeted advertising, the sale of personal data, and profiling in furtherance of decisions that produce legal or similarly significant effects concerning the consumer.[126]

Generally, compared with the right to delete, the right to opt out has a smaller negative impact on the data brokers and their clients because they can still use consumers' data for other purposes.[127] For example, after a consumer opts out of the processing of his data for all purposes listed in the statute, data brokers may still be able to process his data together with other consumers' data to offer marketing analytics products that involve only aggregated data insights. This type of data processing falls outside the scope of the limitations imposed by the statutes because an aggregated

---

124 *See* Vladeck, *supra* note 6, at 509.

125 *See* CAL. CIV. CODE § 1798.120(a) (2024); IOWA CODE § 715D.3(1)(d) (2023); TENN. CODE ANN. § 47-18-3203(a)(2)(F).

126 *See generally* VA. CODE ANN. § 59.1-577(A)(5); COLO. REV. STAT. § 6-1-1306(1)(a); UTAH CODE ANN. § 13-61-201(4); CONN. GEN. STAT. ANN. § 4(a)(5); IND. CODE ANN. § 24-15; MONT. CODE ANN. § 30-14; TEX. BUS. & COM. CODE § 541.051(b)(5); TENN. CODE ANN. § 47-18-3304(a)(2)(E); OR. REV. STAT. § 646A.574(1)(d); DEL. CODE tit. 6, § 12D-104(a)(6).

127 *See* FED. TRADE COMM'N, *supra* note 1, at iii, vi.

data report is not "personal data" and cannot be used against any particular consumer for targeted advertising or profiling.[128]

However, the right to *opt out* of the processing of one's data may cause similar negative consequences as the right to *delete* may cause, including less accurate derived information and more consumers being falsely denied benefits and interests. If data brokers cannot process some consumers' data for targeted marketing purposes, they will not be able to use those consumers' data to train predictive models that aim to produce information derived to facilitate targeted marketing, even though they will not send any targeted advertising materials to those consumers who have opted out.[129] In contrast, allowing consumers to opt out of the selling and sharing of their information will not lead to this problem because data brokers will still be able to process all consumers' data and use the data to train predictive models.[130] As a result of this method, the input data will not be skewed, making the predictive models more accurate.[131] In the meantime, data brokers are not violating the law by "selling or sharing" consumers' data because data brokers do not make predictions about the opted-out consumers or share those consumers' specific information with any other parties.[132] This method will have a minimal impact on the opted-out consumers because their data will not be used against them, and they will not receive any targeted advertising materials, just as if they have opted out of the processing of their data.[133] Therefore, providing consumers with the right to opt out of the selling or sharing of their information rather than the processing of their data could be a better solution from a marketing perspective because it may strike a balance between consumers' privacy rights and the values that businesses may gain from the power of data.

E. Exempting Deidentified and Aggregated Data

The data privacy laws of the thirteen states' exempt deidentified or aggregated data from the limitations discussed above in Part III, subsections A through D. However, the scope of their exemptions varies.[134]

---

128 *See* Kuempel, *supra* note 9, at 218–21.

129 *See* Schmitz, *supra* note 5, at 1427.

130 *See* FED. TRADE COMM'N, *supra* note 1, at 19 (explaining that consumers' data can be used to train predictive models "to apply to other consumers").

131 *See* Vladeck, *supra* note 6, at 495.

132 *See* FED. TRADE COMM'N, *supra* note 1, at 42–43.

133 *See id.*

134 *See generally* CAL. CIV. CODE § 1798.140(m); VA. CODE ANN. §59.1-581; COLO. REV. STAT. § 6-1-1307; UTAH CODE ANN. § 13-61-303; CONN. GEN. STAT. ANN. § 42-523; IOWA CODE § 715D.6; IND. CODE ANN. § 24-15-7-1–3; TENN. CODE ANN. § 47-18-3208; MONT. CODE ANN. § 30-14-2815; TEX. BUS. & COM. CODE § 541.106; FLA. STAT. ANN. § 501.714; OR. REV. STAT. § 646A.583; DEL. CODE tit. 6, § 12D-109-110.

Some types of data processing and disclosure performed by data brokers fall within the exemptions.[135]

First, except Oregon, the remaining twelve states exempt "deidentified" or "pseudonymous" data from at least part of the imposed limitations in a manner that gives data brokers more freedom to do their business.[136] "Deidentified" and "pseudonymous" data refers to personal information that cannot be attributed to a specific individual without the use of additional information, provided that such additional information is kept separately to ensure that the personal data is not attributed to an identified or identifiable individual.[137] This exemption may allow data brokers to analyze consumers' behavior data without personally identifiable information even after the consumer requests to delete his data or opts out of the processing of his data as long as the behavior data is stored separately from the information that can be used to attribute the behavior data to a specific individual. Data brokers can store consumers' purchasing data by transactions after removing personally identifiable information and then perform "market basket analysis" using the Apriori algorithm and the Association Rules, which identify products that are frequently purchased together in one transaction.[138] For instance, this analysis can tell businesses that "if a customer purchases bread, they are also likely to purchase butter."[139] Such information is useful to retailers even if the data is not linked to any specific consumer because they can use it to improve the arrangement of items in physical stores and recommend targeted items to online shoppers based on the items already in their shopping cart.

Second, three states exempt "aggregated" data from some of the imposed limitations.[140] "Aggregated" data refers to information related to

---

135 *See* COLO. REV. STAT. § 6-1-1307; UTAH CODE ANN. § 13-61-303.

136 *See generally* CAL. CIV. CODE § 1798.145(a); VA. CODE ANN. §59.1-581(D); COLO. REV. STAT. § 6-1-1307(3); UTAH CODE ANN. § 13-61-303(2); CONN. GEN. STAT. ANN. §42-523(d); IOWA CODE § 715D.6(3); IND. CODE ANN. § 24-15-7-1, 24-15-7-2; TENN. CODE ANN. § 47-18-3304(a)(4), 47-18-3207(c); MONT. CODE ANN. § 30-14-2815(4); TEX. BUS. & COM. CODE § 541.106(c); FLA. STAT. ANN. § 501.714(3); OR. REV. STAT. § 646A.583; DEL. CODE tit. 6, § 12D-109(c).

137 *See generally* CAL. CIV. CODE § 1798.140(m); VA. CODE ANN. § 59.1-575; COLO. REV. STAT. § 6-1-1303(22); UTAH CODE ANN. § 13-61-101(28); CONN. GEN. STAT. ANN. § 42-515(16); IOWA CODE § 715D.1(23); IND. CODE ANN. § 24-15; TENN. CODE ANN. §§ 47-18-3302(11), (23); MONT. CODE ANN. § 30-14-2802(11); TEX. BUS. & COM. CODE § 541.001(26); FLA. STAT. ANN. § 501.702(13); DEL. CODE tit. 6, § 12D-102(27).

138 *See* Iqra Bismi, *How to Perform Market Basket Analysis Using Apriori Algorithm and Association Rules*, MEDIUM (Jan. 22, 2023), https://medium.com/@iqra.bismi/how-to-perform-market-basket-analysis-using-apriori-algorithm-and-association-rules-3f6ba61d6e4b [https://perma.cc/DLB9-VQEW]. The Apriori algorithm identifies frequent item sets purchased in a single transaction, and the Association Rules identify the relationships between the variables in an item set. *See id.*

139 *Id.*

140 *See* CAL. CIV. CODE § 1798.145(a)(1)(F); TENN. CODE ANN. §§ 47-18-3304(a)(2)(c); FLA. STAT. ANN. § 501.714(3).

a group of consumers that is not linked to or linkable to any individual consumer.[141] This exemption primarily applies to data brokers' marketing analytics products that sell data insights to businesses to facilitate the decision-making process of general marketing strategies.[142] Besides, these statutes may allow data brokers to maintain and use aggregated information that has been generated from the data of consumers who have requested to delete their data.[143]

Overall, these exemptions help mitigate some of the negative consequences caused by the imposed limitations discussed in previous subsections because they allow businesses to leverage the power of data even after consumers have exercised their rights to delete their data or opt out of the data processing program.[144]

IV. PROPOSAL FOR REGULATING DERIVED INFORMATION

Current laws are inadequate to regulate derived information provided by data brokers because these laws cannot balance consumers' privacy rights with businesses' gaining of value from derived information. The complexity of derived information production, the unavoidable inaccuracy of predictive models, and the difficulties of defining whether a prediction related to likelihood and tendency is correct make relying on consumers themselves to protect their privacy rights an inadequate solution. Thus, this Note proposes to regulate derived information through specific statutory limitations enforced by government agencies. The key component of this proposal is shifting from human-focused models to behavior-focused models, aiming to predict consumers' behaviors rather than personal characteristics. Here, the data used to train predictive models should be limited to behavioral data without any information related to demographics or personal characteristics. The removal of non-behavioral information can help mitigate discrimination concerns caused by derived information because it ensures that no protected characteristics will be used against consumers.[145] Besides, this method simplifies the logic behind derived information, making consumers more likely to understand the scope of derived information and thus less likely to suffer subjective

---

141 *See generally* CAL. CIV. CODE § 1798.140(b); FLA. STAT. ANN. § 501.702(2).

142 *See* FED. TRADE COMM'N, *supra* note 1, at 3.

143 *See* CAL. CIV. CODE §§ 1798.140(b), 1798.145(a)(1)(F) (stating that consumers may request to delete personal information, and personal information does not fall within the statutory definition of aggregate data); *see also* FLA. STAT. ANN. §§ 501.702(2), 501.705(2)(c), 501.714(3).

144 *See* FED. TRADE COMM'N, *supra* note 1, at 43.

145 *See* Schmitz, *supra* note 5, at 1454 (arguing that data-based classifications of consumers may heighten discrimination when based on sensitive, identifiable data); *see generally* FED. TRADE COMM'N, *supra* note 1, at B-3 to B-6 (listing examples of demographic and sensitive consumer data elements).

privacy harm.[146] Given the strict restrictions on the predictive models allowed to apply, some other limitations that impose significant burdens on data brokers and businesses while not providing significant protections to consumers' privacy can be lessened, such as the restrictions on sensitive information, the right to access and correct derived information, and the right to opt out of data processing.

A. Shifting from Human-Focused Models to Behavior-Focused Models

Much of the derived information that the Federal Trade Commission considered problematic relates to demographics or other characteristics of consumers themselves, such as "Rural Everlasting," "Thrifty Elders," and "Urban Scramble."[147] This type of derived information comes from sensitive data or discloses sensitive information that was previously unavailable to data brokers' clients.[148] Such derived information can be used in a discriminatory manner. For example, "Thrifty Elders" may be categorized as consumers that are unlikely to spend a lot on electronic devices, making them receive less favorable deals than other consumers. Even derived information that seems unrelated to protected characteristics on its face may have factored in the data related to protected groups.[149] For instance, a score indicating consumers' likelihood of purchasing a product may have taken consumers' age and gender into account.[150]

To deal with this issue, this article proposes to shift from human-focused models predicting the personal characteristics of consumers to behavior-focused models predicting consumers' behavior, including purchase behavior, usage behavior, and consumer loyalty.[151] For example, it would be permissible for data brokers to predict that a consumer will likely buy infant clothes or prefer a particular brand of infant clothes, but not that the consumer is pregnant.[152] All predictions about consumers, including but not limited to demographics, geolocation, financial status, and health condition, should be prohibited even if such predictions are based on nonsensitive and innocuous data.

---

146 *See* Calo, *supra* note 30, at 1134; *see also* Schmitz, *supra* note 5, at 1414.

147 *See* FED. TRADE COMM'N, *supra* note 1, at 20.

148 *Id.*

149 *See Id.*, at 20; Schmitz, *supra* note 5, at 1415; Vladick, *supra* note 6, at 513.

150 *See* FED. TRADE COMM', *supra* note 1, at iii.

151 *See generally* Hemant Warudkar, *How to Analyze and Predict the Behavior of Consumers*, EXPRESS ANALYTICS (Aug. 10, 2021), https://www.expressanalytics.com/blog/how-to-analyze-and-predict-the-behavior-of-consumers [https://perma.cc/ALM5-X62K] (describing the use of consumer behavioral data to predict future consumer behavior and purchases).

152 *See* Anita Ramasastry, *Should Target Tell Your Loved Ones You Are Pregnant, or Should You*? *The Perils of Consumer Data Aggregation, Including Loss of Privacy*, VERDICT (Feb. 28, 2012), https://verdict.justia.com/2012/02/28/should-target-tell-your-loved-ones-you-are-pregnant-or-should-you [https://perma.cc/QR8D-759K].

This method can solve several problems discussed above while allowing data brokers and advertisers to leverage the power of data. First, the logic of predictive models and derived information is simplified, making consumers more likely to fully understand the content and scope of derived information and thus improving transparency. The predictions will only include the likelihood that a consumer will perform certain behaviors in a specific period, such as purchasing a type of item, preferring a particular product feature over another, or spending more than a specific amount of money in a store. Compared with explaining derived information about personal characteristics that makes no sense on its face and needs to be further processed before applying to targeted marketing, explaining the prediction of behaviors is much easier because the meaning of the predictions and how the predictions will be used are straightforward on their faces.[153]

Second, this method reduces the importance of consumers' right to access and correct their data. Consumers can fully understand the scope of derived information from the explanation in the notice, and obtaining specific numeric predictions regarding their future behavior will not help them better understand the information. Also, since the derived information is all about likelihood or probability, consumers do not need to and cannot "correct" it. Thus, the deficiencies in consumers' right to access and correct their data caused by derived information will no longer be a significant problem.

Third, derived information will not provide any previously unavailable information to advertisers that can be used to discriminate against consumers because no prediction will disclose any protected characteristics or sensitive information.

This restriction will not have a significant negative impact on advertisers because, ultimately, their goal is to know which consumers will likely buy their products or how consumers will react to a particular advertisement.[154] Information about consumers' personal characteristics is traditionally used to achieve that goal as well. However, advertisers can also use the predicted probabilities of specific behaviors linked to a consumer to customize their marketing strategies.[155] For example, if an advertiser knows that a consumer has a seventy percent likelihood of

---

153 *See* ME, *Target: You Can't Hide That Baby Bump from Us*, HARV. DIGIT. PLATFORM (Apr. 9, 2018), https://d3.harvard.edu/platform-digit/submission/target-you-cant-hide-that-baby-bump-from-us/ [https://perma.cc/CNK9-QELM].

154 *See* Gary Drenik, *Predicting Consumer Behavior: Do Retailers Know What You'll Buy?*, FORBES (Sept. 12, 2019), https://www.forbes.com/sites/forbesinsights/2019/09/12/predicting-consumer-behavior-do-retailers-know-what-youll-buy/?sh=4423ba316c1e [https://perma.cc/F5EU-K7F6].

155 *See* Qi Zhao & Yi Zhang, *Have Your Cake and Eat It Too! Preserving Privacy while Achieving High Behavioral Targeting Performance*, *in* PROCEEDINGS OF THE CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 1 (2012).

purchasing high-end products, they could emphasize quality over price in the messages sent to that consumer. In contrast, derived information describing consumers' personal characteristics, such as "Thrifty Elders" and "Urban Scramble," does not directly indicate whether a consumer is likely to respond to any type of marketing strategy, which means that advertisers must perform further analysis before applying such derived information in advertising campaigns. Therefore, the absence of personal characteristic data will not have a significant impact on their targeted marketing efficiency.

B. Limiting Input Variables of Predictive Models to Behavioral Data

Because of the unavoidable stereotypes created by predictive models, only removing all protected characteristics from consideration can ensure that the use of derived information in targeted marketing will not lead to discrimination.[156] Nevertheless, prohibiting the use of certain types of factual data in building predictive models is inadequate because algorithms can predict such data based on nonsensitive information.[157] Also, as discussed in subsection C of Part III, forbidding data brokers merely from using predicted sensitive information or data provided by certain sources will not be practical because of the difficulties defining "predicted sensitive information" and "data that may be used to infer sensitive information."[158]

Thus, like the predictions made by models, the input variables used to train the models should also be limited to behavioral data, and no information related to consumers' demographics or personal characteristics can be used. Nevertheless, since all predictions made by models are limited to the likelihood of certain behaviors, derived information can be used as input variables to train the models to make other predictions.[159] For example, the prediction of a consumer's lifetime value in a category of products could be an input variable in a model predicting the consumer's likelihood of purchasing a particular product in that category.[160]

While this method will have some negative impacts on data brokers and advertisers, its benefits outweigh these drawbacks. First, this method offers a clearer standard than the "sensitive information" standard used in existing laws and does not depend on consumers to understand the

---

156 *See* Vladeck, *supra* note 6, at 494–95.

157 *See* Solove, *supra* note 123, at 1084.

158 *See generally* discussion *supra* Part III.C.

159 *See* Rainer Mühlhoff, *Predictive Privacy: Towards an Applied Ethics of Data Analytics*, 23 ETHICS & INFO. TECH. 675, 676–78 (2021).

160 Consumers' lifetime value is a business metric used to determine the amount of money consumers will spend on particular products or services over time. Monique Danao, *What Is Customer Lifetime Value (CLV)*, FORBES (Jun. 14, 2024), https://www.forbes.com/advisor/business/customer-lifetime-value/ [https://per ma.cc/9DTZ-EDSJ].

meaning or the impact of data and to make informed decisions. Second, this method helps avoid discrimination because no potential discriminatory factors will be considered in building the predictive model. Thus, even if the application of the derived information causes disparate impacts on consumers, due to the lack of intent, data brokers and advertisers may still win in courts.[161]

Admittedly, the removal of personal characteristics data from predictive models may lead to lower prediction accuracy, but the extent of the decrease will be acceptable because consumers' behavior data combined with advanced techniques in big data analytics will allow data brokers to build sufficiently accurate predictive models.[162] For example, a model predicting consumers' purchase intent by using anonymous clickstream data achieved an accuracy rate of over eighty percent.[163] Another potential problem of this method is that data brokers' registration targeting products will be influenced because new users do not have any behavior data that can be used to provide customized advertising to them. However, data brokers can still provide aggregated data insights regarding new users' preferences and behaviors to facilitate advertisers in deciding the advertising strategies for new users.[164] Therefore, this method will bring more benefits than drawbacks.

C. Allowing the Prediction of Derived Information Involving Sensitive Features

Since restraining the predictions and the input variables of predictive models to only behavioral data has already removed the discriminatory intent from derived information, predictions involving sensitive features should be permissible. For example, data brokers can predict that a consumer "will likely buy diabetes-related products" but not that a consumer "has diabetes." While the two predictions may have the same effect on advertising strategies, the first prediction will not lead to the harm that may be caused by the second prediction. Compared with understanding how his health conditions are predicted by his purchase history, it is intuitive for a consumer to contemplate that data brokers will know that he will buy products in a particular category after he has

---

161 *See* Vladeck, *supra* note 6, at 513–14.

162 *See* discussion *supra* Part III.C.; *see generally* Christina Vasilopoulou, Leonidas Theodorakopoulos & Ioanna Giannoukou, *Big Data and Consumer Behavior: The Power and Pitfalls of Analytics in the Digital Age*, 45 TECHNIUM SOC. SCIS. J. 469, 469 (2023) ("[B]usinesses can use big data analysis to optimize the customer journey, gain insight into consumer preferences, and create personalized experiences.").

163 Zhanming Wen, Weizhen Lin & Hongwei Liu, *Machine-Learning-Based Approach for Anonymous Online Customer Purchase Intentions Using Clickstream Data*, SYSTEMS, May 18, 2023, at 9–10.

164 *See* discussion *supra* Part III.E.

done so several times, especially after receiving notice regarding how behavior prediction models work. Thus, consumers will be less likely to perceive subjective privacy harm when they receive targeted advertising materials involving sensitive features that are relevant to their previous purchases or other behaviors. Therefore, derived information involving sensitive features should be permitted because it is not generated from any prohibited data and is useful to businesses selling products related to sensitive features, which may also benefit consumers by allowing them to find those products more easily.

D. Lowering Consumers' Control Power over Derived Information

Since the strict limitations on predictive models offer additional protection against discriminatory effects and invasion of consumers' privacy rights, some of the consumers' control power over derived information conferred by current laws can be lowered to counteract the negative impacts on data brokers and advertisers. First, as discussed in subsection A of this Part of the Note, consumers' rights to access and correct derived information are no longer necessary and can be removed.

Second, the right to delete all data and to opt out of the processing of all personal data may result in biased and inaccurate predictive models, imposing negative impacts on data brokers, advertisers, and remaining consumers. Thus, the right to opt out should be limited to the selling and sharing of personal information, including both factual data and derived data. Also, after a consumer requests the deletion of his data, the data broker should be prohibited from making predictions about that consumer and from selling or sharing that consumer's data. However, the data broker should be allowed to use that consumer's data to train predictive models or, at least, to perform analysis aimed at normalizing the skewed input data used in predictive models. This method helps avoid the inaccuracies in predictive models caused by the removal of some consumers' data while simultaneously offering consumers an opportunity to forbid a data broker from using their information against themselves because consumers will not perceive any impacts of the data so long as data brokers cannot make predictions about those consumers or disclose their data to others.

Last, deidentified and aggregated data should be exempted from all limitations discussed in Part IV as long as they do not contain consumers' demographics or personal characteristics. On the one hand, as discussed in Part III, such information can provide valuable insights to advertisers without harming consumers directly because the data cannot be traced back to any specific consumer.[165] On the other hand, if the deidentified or aggregated data links demographics or other personal characteristics to

---

[165] *See* discussion *supra* Part III.E.

behavior data, it will render the limitations on models and input variables useless. For example, using deidentified data containing demographics and purchase data to build predictive models to produce derived information has the same effect as using its identified counterparts because no identification information is necessary for the predictive models. Also, if the aggregated data shows that seventy percent of the purchasers of a particular product are male, it may have the same effect as a model predicting that males are likely to purchase the product while females are not. It may then cause the product's advertiser to offer a different discount based on consumers' gender, which constitutes a simplified and inaccurate human-based model. Therefore, the deidentified and aggregated data exemption should be limited to behavior data.

CONCLUSION

The application of derived information for marketing purposes, fueled by the technological advances of big data analytics, brings significant benefits to businesses and consumers, but it also raises serious privacy concerns. The federal and state governments have enacted various laws during the past few years trying to address this problem, but the complexity and other unique features of derived information make existing laws inadequate.[166] Therefore, this Note proposes a different regulatory framework focused on limiting the types of predictive algorithms and input variables that can be used to produce derived information for marketing purposes while relying less on consumers to make informed decisions to protect their privacy rights. In the meantime, to counteract the negative effects on advertisers imposed by the strict limitations on the predictive models, other restrictions on consumers' rights to access, correct, delete, and opt out should be lowered. This method can better strike a balance between consumers' privacy rights and businesses' marketing efficiency.

166 *See generally* discussion *supra* Part III.