Big Data Analytics Privacy Law Considerations

A Practical Guidance® Practice Note by
Kirk Nahra, Arianna Evers, Ali Jessani, and Genesis Ruano, Wilmer Cutler Pickering Hale

Kirk Nahra
Wilmer Cutler Pickering Hale and Dorr LLP

Arianna Evers
Wilmer Cutler Pickering Hale and Dorr LLP

Ali Jessani
Wilmer Cutler Pickering Hale and Dorr LLP

Genesis Ruano
Wilmer Cutler Pickering Hale and Dorr LLP

This practice note is intended to give privacy practitioners a framework for thinking about the legal issues surrounding Big Data, such as those relating to privacy, data security, and anti-discrimination, and for evaluating potential legal risks, including those related to compliance and consumer protection issues. This practice note approaches Big Data from a U.S. perspective. Increasingly, Big Data will encompass consumer data relating to individuals outside of the United States, in which case other countries' privacy laws will need to be considered for potential applicability, as will any laws governing the transfer of personal data from another country into the United States.

For more related information, see Internet of Things Key Legal Issues and Data Breach Planning and Management.

Big Data—What Is It?

Big Data analytics is the collection and analysis of large and varied data sets (both structured and unstructured) to discover or infer patterns, trends, correlations, and preferences and can be used to make more accurate decisions. Big Data analytics is made possible by the collection of vast amounts of data from a variety of sources, the decreasing cost of obtaining this data, and new technologies and methodologies to analyze data to draw connections and make inferences and predictions.

Big Data analytics is driving innovation across all industries, and there are many benefits to be gained from its analysis. However, there are significant and very real concerns about the risks posed by the use of Big Data. These include the potential for consumer harm, including by perpetuating existing disparities or excluding consumers from receiving the benefits of Big Data. As noted by the Federal Trade Commission (FTC) in its 2016 Report "Big Data: A Tool for Inclusion or Exclusion?," the challenge for organizations is not whether they should use Big Data, but "how organizations can use Big Data in a way that benefits them and society, by minimizing legal and ethical risks."

As described below, the legal landscape surrounding Big Data analytics is uncertain. There is no comprehensive federal privacy law in the United States governing its use, and therefore practitioners must consider an array of different privacy laws and guidance and their potential application to a particular use of Big Data.

Benefits of Big Data

Big Data offers a number of benefits across different industries and practice areas. For example:

- **Healthcare outcomes.** Big Data analytics can be used to predict critical healthcare-related information, develop treatments in areas without specialty providers, and detect and diagnose disease.
- **Business efficiency.** Real-time Big Data analytics can help organizations identify and react to problems in near real-time and be used to analyze sales (and returns) to make better decisions about new products and services, feature enhancements, discontinuations, and other changes.
- **Consumer preferences and personalization.** Big Data allows businesses to provide better services to customers based on their individual needs and changing preferences.
- **Security/fraud.** Big Data can help avoid security threats and fraud by allowing organizations to detect anomalies in their data or on their networks.

Risks of Big Data

Big Data also opens individuals and organizations to significant risks.

Privacy

Privacy is a key concern for Big Data analytics. Many Big Data sets include consumer data and traditional methods used to protect privacy rights, like de-identification or exclusion, may limit the accuracy or usefulness of the analysis. In addition, technologies that rely upon Big Data, such as Internet of Things (IoT) devices, have invaded areas that were historically private and also generate a large amount of sensitive information. For example, smart home devices capture large amounts of information about our day-to-day activities in our most private spaces. For more information regarding IoT, see Internet of Things Key Legal Issues. Some services may use Big Data to personalize offerings in ways that are neither disclosed to nor approved by the user. There is also the risk of large-scale data breaches when data used for analysis is subject to unauthorized access or exfiltration and is used for malicious purposes.

Transparency

Transparency is also a significant challenge in the use of Big Data. Consumers may want to know what data has been collected about them and how it will be used. Many federal and state laws that regulate privacy focus on the need for user consent, notification, and opportunities to opt-out of data collection or use (e.g., Federal Trade Commission Act of 1914 (15 U.S.C. §§ 41–58, as amended) (the FTC Act or the Act) and the California Consumer Privacy Act). Fraudulent or misrepresented use of personal information and consumer data is the basis for a significant portion of FTC enforcement actions related to privacy.

Discrimination and Fraud

The rising use of Big Data analytics can also result in discriminatory hiring and lending practices, among other insidious forms of bias. Big Data analytics can be used to predict individuals' personal, sensitive characteristics, including religion, ethnicity, sexual orientation, and political affiliation, that could be used, in turn, to make decisions that violate consumer protection and equal opportunity laws. This information is often inferred from more traditional, less sensitive, data points (such as addresses, birthdays, and home ownership information). The conclusions derived may also be used by businesses to make dubious and misleading offers or otherwise promote scams to vulnerable individuals, including senior citizens. Incorrect predictions from Big Data can preclude otherwise deserving consumers from credit offers, educational opportunities, and other things, which can have the effect of perpetuating existing disparities.

Errors

Big Data analysis is not error-free and as such, important decisions might be based on false or misleading data points. When poor quality data enters the complex system of Big Data analytics, there can be significant inaccuracies that result in revenue loss, major process inefficiencies, bias or discriminatory effects, and the failure to comply with applicable industry and government laws and regulations.

Data Protection/Security

There is also the risk of unauthorized access or acquisition of large data sets that contain vast amounts of personal information. Attorneys working in Big Data need to understand how to implement reasonable security practices to protect the information, and how to evaluate legal and contractual obligations if there is a data breach.

Relationship to Artificial Intelligence and Internet of Things

Big Data analytics is closely related to Artificial Intelligence (AI) and IoT devices. AI refers to "machines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment, and intention." Darrell M. West, What is Artificial Intelligence?, Brookings Institute (Oct. 4, 2018). A close cousin of AI is the idea of "machine learning," which refers to an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

Big Data is often used to facilitate the use of AI and machine learning, as machines are trained through analysis of large data sets. AI and machine learning techniques can also be used to identify patterns and trends in Big Data sets. These findings can have broad application in healthcare, business, and other industries. One example is the development of generative AI, which refers to algorithms that, after training on massive Big Data sets, can create new outputs, including text, audio, images, or video. In 2022, OpenAI—a generative AI research and development company—launched ChatGPT a chat bot capable of providing answers to queries similar to a web browser and writing a history paper. In just two months, ChatGPT reached 100 million monthly active users—the fastest growth of a consumer application in history.

AI applications facilitated by Big Data sets, such as generative AI, can pose additional legal risks. For example, a chatbot does not limit inputs from users, but it learns from those inputs, this can result in the dissemination of false information. Further, chatbots can be used to quickly generate convincing false information, such as news articles. As such, chatbots can create liability for owners since it can facilitate the creation and dissemination of misinformation. Further, chatbots, like ChatGPT can facilitate ethical and contractual violations if industries that have such obligations to clients (e.g., lawyers, doctors, marketing agencies, etc.) use chatbots, those industries must be aware that inputs are not kept confidential.

IoT is the interconnection and networking of web-connected devices, which are embedded with software, electronics, microchips, sensors, and other forms of technology, that collect, use, exchange, store, transmit, and analyze data. For example, think of a smart fridge or home assistant. Big Data analytics relies heavily on the information gathered, stored, and transmitted by IoT devices (such as buildings, smart devices, appliances, and vehicles, among others).

For more information regarding IoT, see Internet of Things Key Legal Issues.

Legal Landscape—Overview

Entities that want to use Big Data face considerable uncertainty over what legal standards apply and how. There are no comprehensive state or federal privacy laws covering Big Data, so practitioners must look to multiple laws to determine whether or not they apply to the issue at hand.

The FTC Act, under which the FTC can regulate unfair and deceptive trade practices, may cover some Big Data practices. State unfair and deceptive acts and practices laws provide similar enforcement powers to state attorneys general and sometimes private litigants. In the absence of a comprehensive regulatory framework, regulators like the FTC have issued best practices for organizations to follow.

Although there is not yet a comprehensive privacy law in the United States that applies to Big Data, there are three states with comprehensive privacy laws that may affect an entity's Big Data practices if those practices involve personal information. In addition, there are several federal laws that may cover Big Data practices depending on the industry or type of data at issue.

The following sections of this practice note highlight some of the privacy laws that practitioners should consider when thinking about Big Data issues.

Federal Trade Commission

The FTC has jurisdiction over most for-profit organizations and individuals doing business in the United States, other than those in the telecommunications, financial, and transportation industries, which are primarily regulated by other federal agencies. (Note that nonprofits are generally excluded from the FTC's jurisdiction.) The FTC's authority under Section 5 of the FTC Act extends to "unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce," except for select industries that are regulated by other federal laws and agencies. 15 U.S.C. § 45(a)(1). The FTC can bring enforcement actions relating to Big Data under its Section 5 authority or one of the many statutes or rules that it is responsible for enforcing. The FTC also issues guidance that indicates how the FTC views certain issues, and practices inconsistent with that guidance have the potential to result in corrective action by the Commission under Section 5 if those practices are found by the Commission after an investigation to be unfair or deceptive.

General Section 5 Authority

The FTC Act was established to regulate questionable business practices and protect consumers. Specifically, Section 5 of the FTC Act prohibits unfair or deceptive acts and practices in commerce, which includes consumer privacy violations and engaging in improper data collection, use, and disclosure practices (including where those practices involve Big Data). 15 U.S.C. § 45. Section 5 is also routinely applied to penalize organizations that do not have reasonable data security practices. The FTC can bring enforcement actions for Section 5 violations.

Practitioners responsible for the collection, storage, use, disclosure, or other processing of Big Data should ensure that those activities do not violate Section 5's prohibition on unfair or deceptive acts or practices. A three-part test is used to determine whether an act or practice is deceptive:

- The representation, omission, or practice must mislead or be likely to mislead the consumer.
- The consumer's interpretation of the representation,

omission, or practice must be reasonable under the circumstances.

- The misleading representation, omission, or practice must be material.

See FTC Act Policy Statement on Deceptive Acts and Practices.

This means that practitioners should be careful about statements that could be seen as causing a consumer to be misled in any material way. Organizations should act in a manner that is consistent with the promises made to consumers about the collection, use, storage, or dissemination of personal information. In addition, organizations should seek affirmative express consent where the data involved is particularly sensitive (e.g., health or financial information) and where they seek to use data in a manner that is materially different from the purpose for which it was originally collected.

Unfair acts or practices are those that:

- Cause or are likely to cause substantial injury (usually monetary) to consumers
- Cannot be reasonably avoided by consumers –and–
- Are not outweighed by countervailing benefits to consumers or competition

Public policy may also be considered in the analysis of whether a particular act or practice is unfair. For example, the FTC typically regulates data security under the unfairness prong of its authority and has (through consent orders and guidance) developed standards for what constitutes reasonable security practices.

FTC Big Data Practices Settlements

The FTC has reached a number of settlements involving organizations' Big Data practices. The conduct at issue has involved both violations of Section 5 and other statutes or rules, such as the Children's Online Privacy Protection Act (COPPA), or the Health Breach Notification Rule (HBNR), over which the FTC has jurisdiction.

For example:

In 2023, the FTC reached a settlement with GoodRx – a digital health platform – for allegedly sharing users' personal health information with third parties without properly disclosing their data practices or obtaining users' affirmative consent, as well as for failing to maintain adequate policies or procedures to protect users' personal health information. GoodRx operates a digital health platform which offers prescription drug discounts, telehealth visits, and other health services. Consumers can access discounts by providing GoodRx with personal information such as the name of a particular medication, dosage, and location information. The FTC Complaint alleges that GoodRx represented to consumers that they would never disclose personal health information to advertisers, but they deceived consumers by sharing their sensitive health information for targeted advertising purposes in violation of these representations. Further, according to the complaint, GoodRx claimed that consumer information would be disclosed only for limited purposes. However, GoodRx allegedly did not limit third parties' use of data, allowing for third parties to use consumer data for their own internal purposes, including for research and development or to improve their advertising services. The FTC found that together, these acts were unfair to the consumer since the consumer had no notice of the practices nor could they reasonably avoid these uses of their information. Notably, the complaint alleged that since GoodRx disclosed consumer health information without affirmative express consent, the sharing of such information was an unauthorized disclosure, which requires notification under the HBNR.

Per the proposed consent order, GoodRx agreed to pay $1.5 million, to implement additional policies and procedures with regards to its data privacy practices, as well as adhere to a number of substantive limits on GoodRx's data practices. Specifically, GoodRx is required to implement and document a comprehensive privacy program. Further, the order prohibits GoodRx from disclosing health information for targeted advertising purposes and from sharing health information for any other purpose without affirmative express consent or notice.

The GoodRx decision demonstrates the FTC's concern companies that collect and use large amounts of consumer information, especially when that personal data is sensitive and when the uses include online behavioral advertising. This concern is also highlighted by the FTC's litigation with Kochava, a data broker that, according to the FTC's complaint, collected and sold the precise location data of individuals. The complaint specifically notes that precise geolocation data when associated with advertising IDs—as sold by Kochava— is not by nature anonymized and reveals intimate details about consumer's lives such as sensitive locations visited by the consumer. Further, according to the FTC, Kochava allegedly maintains no technical controls to prevent this and continues to sell such data. According to the FTC, these actions are unfair in that they are likely to cause substantial injury to consumers which consumers cannot reasonably avoid. Big Data companies focused on analyzing specific types of sensitive information should keep the FTC's continued focus in mind.

Also in 2022, the FTC reached a settlement with WW International Inc., formally known as Weight Watchers, regarding allegations that the company collected children's information without parental consent in violation of COPPA. According to the complaint filed by the Department of

Justice (DOJ) on behalf of the FTC, Weight Watchers marketed a weight management and tracking service designed for use by children ages 8–17, named "Kurbo by Weight Watchers" (Kurbo). Until late 2019, children could sign up for the service either by indicating they were a parent signing up for their child or a child over the age of 13. The complaint alleged that Kurbo's "age gating" sign-up process encouraged younger users to falsely claim they were over the age of 13, despite text indicating that children under 13 must sign up through a parent. The complaint also alleged that hundreds of users who signed up for the app claiming to be over the age of 13 later changed their birthdates on their profiles to indicate they were really under 13 and that these users continued to have access to the app until FTC contacted Kurbo. Further, the complaint alleged that Kurbo's information collection notification and storage practices failed to comply with COPPA.

As part of its settlement with the FTC, Weight Watchers and Kurbo were required to delete personal information collected from children, destroy any models or algorithms derived from the data, and pay a $1.5 million penalty. Additionally, the settlement required Weight Watchers and Kurbo to maintain a schedule for retaining children's data for no longer than one year after the last instance in which the user tracks their food intake, weight, or activity level.

In 2021, California-based photo app Everalbum reached a settlement with the FTC regarding allegations that the company misled consumers about its use of facial recognition technology. At issue was a photo storage and organization application—"Ever"—that was directed to consumers and included a face recognition feature. Although Everalbum's practices regarding the face recognition feature varied over time, at least some subset of users had the face recognition feature enabled by default and were unable to turn it off until April 2019. At that time, Everalbum disabled the feature and face recognition for all users until they clicked "yes" on a pop-up message that asked for their permission to use face recognition. Everalbum also used certain photos that it collected from individuals to develop its own facial recognition technology. The FTC also alleged that Everalbum failed to delete the photos of users who had deactivated their accounts despite multiple statements suggesting that account deactivation would delete all associated photos.

As part of its settlement with the FTC, Everalbum was required to delete facial recognition models that were created from its consumer users' photos, even though there were no counts in the complaint directly relating to Everalbum's use of the improperly collected photos to develop its technology. Additionally, the settlement included requirements that Everalbum obtain consent for the app's facial recognition features and mandate the deletion of photos from deactivated accounts. Everalbum is also prohibited from misrepresenting how it collects, uses, discloses, maintains, or deletes personal information, including face embeddings created with the use of facial recognition technology, as well as the extent to which it protects the privacy and security of personal information it collects.

Also in 2021, the FTC settled with Flo Health Inc., a developer of a period and fertility-tracking app used by more than 100 million consumers. According to the FTC's complaint, the Flo Period & Ovulation Tracker was using a number of software development kits (SDKs) from various third-party marketing and analytics firms. These SDKs gathered the unique advertising or device identifiers of users, as well as standard and custom app events, and shared them with the marketing and analytics firms for various purposes. The custom app events were set up by the application's developers to have descriptive titles that conveyed sensitive health information. One example from the complaint was the use of the title "R_PREGNANCY_WEEK_CHOSEN" for when a user entered the week of her pregnancy. According to the FTC, sharing these custom app events—because of what the titles conveyed—not only violated Flo's own privacy policy, but also violated the terms of use of the third parties with whom Flo was sharing this information.

As part of its settlement agreement with the FTC, Flo was required to notify users through its website and by email that Flo had shared an identifying number and personal health information with third parties, as well as about the settlement with the FTC. Additionally, Flo is prohibited from misrepresenting the purposes for which it or entities to whom it discloses data collect, maintain, use, or disclose the data; how much consumers can control these data uses; its compliance with any privacy, security, or compliance program; and how it collects, maintains, uses, discloses, deletes, or protects users' personal information. In addition, Flo must instruct any third party that received users' health information to destroy that data.

The Singapore-based mobile advertising company InMobi paid $950,000 as part of a 2016 settlement over its use of consumer location data. (The penalty was originally $4 million but was reduced to $950,000 as a result of the company's financial condition.) According to the allegations in the FTC's complaint, InMobi allegedly tracked the location of hundreds of millions of users who had either explicitly refused or had never been asked to consent to location tracking and used the location information to target ads in third-party applications. In addition to misleading adult consumers, the company allegedly knowingly tracked the location of children--a violation of the COPPA. The settlement required InMobi to delete collected location data, establish robust privacy practices, and submit to 20 years of third-party privacy assessments.

For more information regarding COPPA, see Children's Online Privacy Protection Act (COPPA) Compliance.

FTC Guidance

The FTC has also issued guidance relevant to Big Data, specifically a staff report on Big Data, as well as guidance on AI, Dark Patterns, and the Safeguards Rule that should be considered by practitioners who work in these areas.

FTC Big Data Report (2016)

In January 2016, the FTC issued a report entitled Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues. The report acknowledges the rapid expansion of Big Data analytics as a natural outgrowth of technological advancement (including the proliferation of computers, smartphones, and IoT devices), and commends its success in guiding new product development, predicting individual preferences, tailoring services and opportunities, and guiding individualized marketing, among other things. The report, which primarily addresses the commercial use of Big Data (i.e., the exploitation of consumer information), acknowledges that while Big Data has the potential to create opportunities for some consumers, it also can serve to deprive many other consumers of such opportunities (especially members of low-income and/or underserved populations). Additionally, it discusses the benefits and risks created using Big Data analytics, as well as the consumer protection and equal opportunity laws that apply to its use and potential exploitation. The report also provides guidance to businesses that use Big Data on how to maximize its benefits, minimize its risks, and maintain compliance with applicable law. It urges caution for businesses that use Big Data to ensure that their uses of Big Data do not result in discriminatory or harmful outcomes. The report does not discuss the collection of Big Data and only touches on security, privacy, notice, and choice in a limited manner.

FTC Guidance on Artificial Intelligence

The FTC primarily regulates businesses' use of AI under three laws: Section 5 of the FTC Act, the Fair Credit Reporting Act, 15 U.S.C. § 1681 (FCRA), and the Equal Credit Opportunity Act, 15 U.S.C. § 1691 et seq. (ECOA). In 2021, the FTC summarized several principles for businesses to follow as they deploy AI tools:

- Use a data set that does not exclude particular populations or groups
- Test algorithms to ensure they are not discriminating based on protected characteristics
- Foster transparency and allow independent evaluations of your AI use
- Don't publish false advertising on product capabilities
- Be truthful about data use
- Ensure that positive uses of AI outweigh harms
- Be prepared to be held accountable.

These principles tracked the FTC's 2020 guidance on AI and algorithms for businesses. That guidance warned that while more and more businesses are deploying AI, the FTC has decades of experience prohibiting unfair, deceptive, or otherwise unlawful use of automated decision-making. The FTC warned businesses that AI tools do not absolve businesses of their responsibilities to issue "adverse action" notices or fulfill other regulatory responsibilities, nor does the use of AI mean businesses can be ignorant of how they are making decisions.

FTC Guidance on Dark Patterns

In September 2022, the FTC released a report entitled "Bringing Dark Patterns to Light." Data that is used for Big Data and AI can be collected in various manners, and some forms of collection, such as through the use of dark patterns, can heighten regulatory risk for institutions that collect and analyze data. This FTC guidance identifies the types of misleading and manipulative interface practices that the agency believes are Dark Patterns and can harm consumers and emphasizes the agency's focus on enforcement in the area. Further, it offers guidance on what data controllers and processors should be paying attention to with respect to how the data they use is collected. Specifically, the report focuses on dark patterns that hide or delay disclosure of material information and design elements that lead to unauthorized charges or obscure or subvert privacy choices, among others. Dark patterns used to subvert privacy choices are of particular importance in the Big Data context. As the report notes design elements that obscure or subvert privacy choices, like using default settings that maximize data collection or make data collection difficult to avoid, can result in illusory choices that steer consumers into unknowingly sharing additional data. Companies should avoid the use of dark patterns as described above because inappropriate data collection methods can lead to increased regulatory risk and taint both the data collected by companies, as well as any models created using that data. For additional guidance, the FTC includes an Appendix to the report that identifies the dark pattern types and their corresponding descriptions.

FTC Guidance on Safeguards Rule

Entities that collect large amounts of personal data need to think carefully about data security, and the FTC is increasingly requiring that companies have certain baseline protections in place. In May 2022, the FTC released a publication entitled "FTC Safeguards Rule: What Your Business Needs to Know." The publication offers guidance to financial institutions on the FTC's revised Safeguards Rule under the Gramm-Leach Bliley Act (GLBA), Pub. L. 106-

102, 113 Stat. 1338. Although the guidance only applies to financial institutions regulated by the FTC, the guidance illustrates the types of safeguards that the FTC expects companies processing sensitive personal information should have in place. For example, the publication notes that a reasonable information security program must include nine elements:

- A qualified individual responsible for the security program
- Periodic risk assessments
- Safeguards to control the risks identified through risk assessments
- Monitoring and testing effectiveness of safeguards on a regular basis
- Train staff regularly on cybersecurity awareness
- Service provider oversight
- Keeping information security program current to safeguard against emerging threats
- Creating a written incident response plan –and–
- Annual reports to boards of governors on security program

This is consistent with the FTC's 2022 consent order with CafePress over allegations that it failed to secure consumers' sensitive personal data, covered up a data breach, and failed to abide by its own representations related to how it uses and discloses consumer data. The consent order required CafePress to develop written information security plans that address the underlying issues identified by the FTC, including security vulnerabilities, inadequate security protocols, procedures for honoring consumer privacy request, among others. In a companion consent order, Residual Pumpkin Entity, LLC, the former owner of CafePress, agreed to pay $500,000 in redress to victims of CafePress data breach.

FTC Rulemaking on Commercial Surveillance and Data Security

On August 11, 2022, the FTC published an Advanced Notice of Proposed Rulemaking (ANPR) to request public comment on the prevalence of "commercial surveillance and data security practices" that harm consumers. This is the first concrete step by the agency to explore using its Section 18 rulemaking under the FTC Act to issue a broad consumer privacy-focused trade regulation rule. Though this will be a lengthy process, some of the topics that the FTC plans to address through this rulemaking will be particularly relevant for companies involved in Big Data, including issues related to algorithmic error and bias, as well as the effectiveness of company notice, transparency, and disclosure. The next steps for this process are for the FTC to issue a Notice of Proposed Rulemaking, which will include the text of the proposed rule.

State-Level Unfair and Deceptive Acts and Practices Laws

Every state has a consumer protection law that prohibits deceptive practices, and many prohibit unfair or unconscionable practices as well. These are referred to as "UDAP" (Unfair and Deceptive Acts and Practices) laws. State attorneys general typically enforce UDAP statutes and can use them to regulate Big Data, much like the FTC uses its Section 5 authority to police Big Data abuses. Such investigations and settlements can be high-profile. For example, nearly every state attorney general participated in a settlement with Equifax over its 2017 data breach. Equifax, one of the three largest credit bureaus (a data broker that essentially trades in Big Data), agreed to pay $600 million to settle allegations that it failed to safeguard the sensitive personal information of almost 150 million people.

California (CCPA/CPRA)

In 2018, the California Consumer Privacy Act (CCPA) (Cal. Civ. Code § 1798.100 et seq.) became the first comprehensive privacy law to be passed in the United States. The law incorporates privacy principles akin to the General Data Protection Regulation (GDPR) in the EU (though with notable differences). It requires businesses that process personal information about California residents and meet certain revenue or data processing thresholds to comply with a number of data privacy requirements. For more information, see California Consumer Privacy Act (CCPA) Overview.

For example, the CCPA gives users more control over the collection of their data by providing them with individual rights, such as the right to know what information a business processes about them (Cal. Civ. Code § 1798.100), the right to delete their information (Cal. Civ. Code § 1798.105), and the right to opt-out of the sale of their personal information (Cal. Civ. Code § 1798.120). The CCPA also requires businesses to provide certain disclosures to California residents, including in their privacy policy and through a "notice at collection," as well as to implement certain contractual provisions with service providers (Cal. Civ. Code §§ 1798.100, 1798.130, and 1798.140(w)).

The CCPA has a broad definition of personal information. Instead of regulating information collected in a certain context (like the way the Health Insurance Portability and Accountability Act of 1996 (Pub. L. 104-191, 110 Stat. 1936) (HIPAA) does for covered entities and the GLBA does for financial institutions), the CCPA regulates all "personal information" that a business collects about California residents (unless the information is otherwise exempted from the scope of the law).

The CCPA defines personal information as "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household" and includes categories of information such as:

- Identifiers
- Commercial information
- Biometric information
- Internet activity information
- Geolocation data

Cal. Civ. Code § 1798.140(v).

Notably, the CCPA categorizes the inferences made by profiling a consumer—inferred by using the categories of information listed above to determine consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes—as personal information. Cal. Civ. Code § 1798.140(v)(1).

This broad definition of personal information has serious implications for companies looking to engage in Big Data analytics as much of the information they process may fall within the scope of the law. To the extent a company analyzing Big Data is subject to the CCPA and is processing personal information under the law, that company would need to provide the appropriate notice to individuals and give them the opportunity to exercise their rights under the law.

While the CCPA has a broad definition of personal information, it also has broad exemptions. For example, the law exempts:

- Personal health information processed in accordance with HIPAA (see Cal. Civ. Code § 1798.146(a)(1))
- Information processed by financial institutions in accordance with the GLBA (see Cal. Civ. Code § 1798.145(e))
- Until January 1, 2023, information processed in employment and business-to-business (B2B) contexts (though these exemptions are somewhat limited) (see Cal. Civ. Code §§ 1798.145(h), 1798.145(n))

These broad exemptions have generally been a trend at the state law level.

The CCPA is enforceable by the California Attorney General (California AG) for privacy and data security-related violations (with fines up to $7,500 per violation). Cal. Civ. Code § 1798.155(b).

The law also includes a narrow private right of action for security incidents that occur as a result of a business's failure to implement and maintain reasonable security procedures and practices. Cal. Civ. Code § 1798.150(a)(1). Note, however, that the definition of personal information in this provision in the law comes from California's breach notice statute instead of the CCPA and is significantly narrower. Cal. Civ. Code § 1798.81.5(d)(1)(A). The fines under the CCPA's private right of action are also limited to $750 per violation. Cal. Civ. Code § 1798.150(a)(1)(A).

The CCPA provides businesses with a 30-day "right to cure" for both California AG enforcement actions and data breach lawsuits. Cal. Civ. Code § 1798.155(a) (California AG enforcement actions); Cal. Civ. Code § 1798.150(b) (private civil actions).

The California AG is responsible for rulemaking under the law and the CCPA points to areas of particular focus for the AG. Specifically, the CCPA authorizes the AG to issue regulations governing access and opt-out rights with respect to businesses' use of automated decision-making technology, including for purposes of profiling. The AG has provided detailed steps that businesses must take in order to comply with the CCPA, for example, requiring businesses' response to access requests to include meaningful information about the logic involved in automated decision-making processes, as well as a description of the likely outcome of the process with respect to the consumer. Cal. Civ. Code § 1798.185.

While the CCPA continues to be in effect and enforcement is ongoing, Californians passed a new privacy law by ballot initiative in November of 2020. The California Privacy Rights Act (CPRA) (see Cal. Civ. Code § 1798.100, Note to 2020 Amendment) both replaces and builds upon the CCPA and brings the law even more in line with the GDPR. Among other notable changes, the law creates a new category of information labeled "sensitive personal information," that is subject to special notice and opt-out requirements under the law. Cal. Civ. Code § 1798.100, Note to 2020 Amendment. The CPRA also creates a new enforcement agency, the California Privacy Protection Agency, that is responsible for rulemaking and enforcement for both the CCPA and CPRA. Cal. Civ. Code § 1798.100, Note to 2020 Amendment. The CPRA goes into effect on January 1, 2023 (and most of the provisions of the law apply to information that a business processes after January 1, 2022). Cal. Civ. Code § 1798.100, Note to 2020 Amendment.

Virginia Consumer Data Protection Act

In March of 2021, Virginia became the second state to pass a comprehensive privacy law. Va. Code Ann. § 59.1-575 et seq. The Consumer Data Protection Act (CDPA) mirrors the CCPA/CPRA and GDPR in many respects, including by:

- Providing Virginia residents with individual rights (Va. Code Ann. § 59.1-577)
- Creating special obligations for businesses that process, including by automated means, "sensitive" data (such as data relating to racial or ethnic origin, genetic or biometric

data, or precise geolocation data) (Va. Code Ann. § 59.1-575)

- Requiring businesses to implement certain contractual provisions with their service providers (Va. Code Ann. § 59.1-579(B))

The CDPA differs from the CCPA/CPRA in meaningful ways, including that the CDPA:

- Has broader exemptions to the law for certain types of information (Va. Code Ann. § 59.1-576)
- Requires businesses to implement certain privacy-by-design principles (such as data minimization and purpose limitation) (Va. Code Ann. § 59.1-578)
- Provides data subjects with the specific right to opt-out of processing of personal data for the purposes of targeted advertising or profiling (Va. Code Ann. § 59.1-577)
- Does not have a private right of action
- Does not provide the Virginia Attorney General with any rulemaking authority

However, like the CCPA/CPRA, the CDPA gives Virginia Attorney General exclusive enforcement authority. Va. Code Ann. § 59.1-584(A). Fines under the CDPA can also be as high as $7,500 per violation. Va. Code Ann. § 59.1-584(C).

The CDPA goes into effect on January 1, 2023.

Colorado Privacy Act

In June of 2021, Colorado became the third state to join the patchwork of laws in the United States when its legislature passed the Colorado Privacy Act (CPA). C.R.S. 6-1-1301 et seq. The majority of the provisions of the CPA go into effect on July 1, 2023.

The CPA includes similar provisions to both the CCPA/CPRA and CDPA in terms of providing, among other similarities:

- Individual rights for consumers, including the right to opt out of profiling by automated means (C.R.S. 6-1-1306)
- Notice requirements for businesses (C.R.S. 6-1-1308(1))
- Special protections for sensitive data (C.R.S. 6-1-1308(7))

Notably, the CPA classifies the processing of personal data for the purposes of profiling where it presents a foreseeable risk of harm to a consumer, as a form of heightened risk processing. Further, it requires controllers that engage in such processing to conduct a data protection assessment. C.R.S. 6-1-1309(2). Unlike California and Virginia's data privacy laws, the CPA provides that, in addition to the Colorado Attorney General, the law is also enforceable by district attorneys. C.R.S. 6-1-1311(1)(a). Additionally, instead of creating separate statutory damages, the CPA is enforceable as unfair trade practice under existing Colorado law. C.R.S. 6-1-1311(1)(c).

Utah Consumer Privacy Act

In March 2022, Utah became the fourth state to pass a comprehensive privacy law. Most provisions of the Utah Consumer Privacy Act (UCPA), Utah Code Ann. § 13-61-101 et seq., effective Dec. 31, 2023.

The UCPA includes similar provisions to the CCPA/ CPRA and CDPA in terms of providing, among other similarities:

- Individual rights for consumers (UCPA § 13-61-201)
- Notice requirements for businesses (UCPA § 13-61-302)
- Special protections for sensitive data (UCPA § 13-61-302 (3))

However, the UCPA most closely mirrors the CDPA, differing from other state privacy laws in meaningful ways, including that the UCPA:

- Provides broad exemptions for entities regulated under certain federal laws, including government entities; covered entities and business associates under HIPAA; information governed by HIPAA; financial institutions and information governed by the GLBA; and personal data regulated by FERPA. (UCPA § 13-61-102)
- Does not provide consumers with the ability to opt-out of processing using a global privacy control (UCPA § 13-61-201)
- Does not provide the Utah Attorney General with any rulemaking authority (UCPA § 13-61-402)

Unlike Virginia's data privacy laws, the UCPA applies to a broader group of businesses, for example, businesses that have $25 million in gross revenue are covered under the UCPA but not the CDPA. Further, in regard to collection of sensitive data, the UCPA requires notice and an opportunity to opt out, while the CDPA requires only consent. Perhaps the largest difference between the similar privacy laws is that the UCPA does not require data protection assessments while the CDPA, like the CPA, require impact assessments when processing sensitive data that represents a heightened risk of harm, such as profiling data.

Connecticut Data Privacy Act

In May 2022, Connecticut became the fifth state to join the patchwork of laws in the United States when its legislature passed the Connecticut Data Privacy Act (CTDPA) of 2022, S.B. No. 6. The majority of the provisions of the CTDPA go into effect on July 1, 2023.The CTDPA includes similar provisions to other existing state privacy laws. For example, the CTDPA creates notice requirements and individual rights for consumers, including the right to opt out of processing for the purposes of targeted advertisements and profiling by automated means. Like the CPA, the CTDPA creates additional requirements for the processing of sensitive data, including profiling data collected by automated means. (CTDPA § 8)

However, the CTDPA most closely mirrors Colorado's privacy laws, differing from other state privacy laws in meaningful ways, including that the CTDPA:

- Permits consumers to designate an authorized agent to opt out on their behalf (CTDPA § 5)
- Allows for enforcement under the state's unfair trade practice under existing Connecticut law (Conn. Gen. Stat. § 42-110b)
- Does not provide any broad-based exemptions for covered entities or business associates regulated under HIPAA; the relevant HIPAA exemption instead applies to protected health information regulated under HIPAA (CTDPA § 3(b))
- Does allow for the creation of a working group that would make recommendations to the legislature as to potential amendments to the law (CTDPA § 12)
- Does not provide the Connecticut Attorney General with any rulemaking authority (CTDPA § 11(a))

Unlike Colorado's data privacy law, the CTDPA creates stricter notice requirements. For example, the CTDPA requires clear and conspicuous links to a webpage for opting out of a sale or targeted advertisement and an active email address for how to contact the controller. Perhaps the most meaningful difference between the Connecticut state privacy law and other state laws is that the CTDPA explicitly carves out payment transaction data from its applicability threshold.

Practitioners advising clients of their various obligations under these state laws should take note of these state law differences when developing compliance programs for their Big Data analytics programs.

Future State Comprehensive Privacy Laws

You should also be aware of the evolving privacy landscape, particularly as it pertains to future state comprehensive privacy laws. In 2023, at least 21 states proposed some version of a comprehensive privacy law. One state, Iowa, had a proposal pass both of its legislative chambers, making it likely to be the sixth state with such a law in place (pending the governor's signature).

Federal and State Sector or Context-Specific Privacy Laws

There are a number of federal and state laws that either regulate specific sectors and therefore may touch on Big Data, or that will apply depending on what data is in the set, or how it is being collected or used. The list below is not exhaustive, but is meant to provide practitioners with an overview of the types of laws that can intersect with Big Data applications.

Financial Services

Below are federal laws in the financial services sector that intersect with Big Data applications.

Gramm-Leach-Bliley Act

The GLBA requires organizations that offer financial services and products to:

- Provide their customers with a detailed account of their information-sharing practices –and–
- Safeguard their customers' sensitive data

The Financial Privacy Rule, promulgated under the GLBA, requires financial institutions to provide each customer with a written privacy notice when establishing a relationship and thereafter on an annual basis and every time the policy is updated. The notice must include, among other things:

- What information is collected
- Where and with whom information is shared
- How much information used
- How information is protected
- Notice to customer that he or she has the right to opt-out of the sharing of personal information with nonaffiliated third parties, subject to certain exceptions

The Safeguards Rule, also promulgated under the GLBA, requires financial institutions to develop, implement, and maintain a comprehensive information security program that consists of the administrative, technical, and physical safeguards the financial institution uses to access, collect, distribute, process, protect, store, use, transmit, dispose of, or otherwise handle customer information. Specifically, banking financial institutions are required by their respective governing agencies to have a written information security program, undertake comprehensive risk assessments, appoint a qualified individual to be responsible for the institution's information security program, regularly conduct independent third-party testing of key controls, protect consumer information through encryption, and much more.

For more information regarding GLBA, see Gramm-Leach-Bliley Act (GLBA) Privacy Requirements.

Fair Credit Reporting Act

The FCRA regulates the practices of consumer reporting agencies (CRAs) that collect and compile consumer information into consumer reports for use by credit grantors, insurance organizations, employers, landlords, and other entities in making eligibility decisions affecting consumers.

CRAs generally include credit bureaus, employment background screening organizations, and other businesses that help organizations make consumer eligibility

determinations (such as tenant screening organizations) for employment, credit, housing, insurance, or other similar decisions.

According to the FTC, the FCRA was enacted to:

- Prevent the misuse of sensitive consumer information by limiting recipients to those who have a legitimate need for it
- Improve the accuracy and integrity of consumer reports
- Promote the efficiency of the nation's banking and consumer credit systems

The FCRA regulates the collection, transmission, and use of private consumer data (including credit information) and serves to protect consumers from the negligent or willful inclusion of inaccurate information in consumer reports (see 15 U.S.C. § 1681 et seq).

It also dictates how CRAs must maintain consumer files, how parties may provide information about consumers to CRAs, how information contained in any reports may be disputed, and how an individual or entity may request and/or use a report (15 U.S.C. §§ 1681e, 1681g, 1681i, 1681m). Specifically, CRAs are obligated to implement reasonable procedures to ensure the accuracy of information contained in their reports, and provide consumers with access to their own information, along with the ability to correct any errors (15 U.S.C. §§ 1681c, 1681c-1, 1681c-2). The FCRA is enforced by the FTC and the Consumer Financial Protection Bureau.

For example:

The FCRA is applicable when a company purchases predictive analytics that contain information not generally included in traditional credit or background checks (including, by way of example, a consumer's social media usage, zip code, and shopping history) to make eligibility decisions. Even though the data sets might not contain credit scores, for example, to the extent they are being used to make eligibility decisions, practitioners will want to carefully consider the FCRA's application to avoid engaging in discriminatory or other fraudulent practices. Practitioners should be aware that whether or not an entity is a CRA depends on a number of factors (see 15 U.S.C. § 1681(a)(f)) (definition of "consumer reporting agency"), and that disclaimers that claim that an organization is not a credit reporting agency, and therefore not subject to the FCRA, have been found insufficient to insulate organizations from FTC enforcement..

The FTC now points to the FCRA as one of the mechanisms to regulate use of AI and algorithms in its guidance on AI and its Big Data report. The FCRA may require that decisions made using AI on credit, housing, or other types of eligibility may need to be supported with an "adverse action" notice, which notifies a customer of their right to check the accuracy of the underlying information.

For more information regarding FCRA, see Fair Credit Reporting Act.

Equal Credit Opportunity Act

The ECOA prohibits credit discrimination on the basis of race, religion, nationality, sex, color, marital status, age, or because an individual receives public assistance. The law applies to any individual who regularly participates in the making of a credit decision (such as banks, credit card organizations, retailers, credit unions, and financial institutions).

To prevail on an ECOA claim, a plaintiff must show either:

- "Disparate treatment" (i.e., when a creditor treats an applicant differently based on a protected characteristic such as race or marital status) –or–
- "Disparate impact" (i.e., when a creditor does not engage in disparate treatment yet otherwise employs practices that have an adverse effect on a protected class)

Disparate treatment is not found to exist where the practice in question exists to further a legitimate business need that cannot reasonably be achieved by means that are ultimately less disparate.

The ECOA requires creditors to:

- Inform a candidate if he or she has been granted or denied credit within 30 days of receiving his or her completed application –and–
- Provide, in sufficient detail, the reason for any denial

Regulation B of the ECOA specifically prohibits creditors from making written or oral statements, in advertising or elsewhere, to potential applicants that would reasonably discourage their application.

Practitioners should urge their clients to avoid exploiting Big Data to make credit eligibility decisions based upon personal, nontraditional background check information to maintain ECOA-compliance and avoid an enforcement action.

Fair Debt Collection Practices Act

The FDCPA (15 U.S.C. § 1692 et seq.) establishes legal protections against abusive debt collection practices. It amends the Consumer Credit Protection Act by becoming Title VIII of that act.

To promote fair debt collection and limit abusive collection practices, the FDCPA sets forth guidelines under which debt collectors may conduct business, establishes consumer rights with respect to debt collection and privacy rights, and provides penalties for violators. The FDCPA is enforced by the FTC.

For more information regarding the FDCPA, see Fair Debt Collection Practices Act.

New York State Department of Financial Services Cybersecurity Regulations

The New York State Department of Financial Services (DFS) is responsible for regulating financial services and products, including those businesses that are subject to the New York insurance, banking, and financial services laws. DFS recognizes that the financial service industry generates a large amount of data. This has enabled the industry to take advantage of data analytics, including the use of machine learning, AI, and data aggregation. Without proper cybersecurity protocols in place, a cybersecurity incident that affects a financial service institution has the potential to compromise large amounts of data. In March 2017, DFS adopted comprehensive cybersecurity regulations for financial institutions that fall under its purview, the Cybersecurity Regulations. The purpose of the Cybersecurity Regulations is to provide businesses with clear guidance to ensure sufficient compliance, especially in times of cybersecurity crises.

Among other requirements, the Cybersecurity Regulations require covered businesses to:

- Conduct and document annual risk assessments, penetration testing, and vulnerability assessments
- Establish written cybersecurity policies based on risk assessments which must addresses information security, data governance, and classification
- Conduct asset inventory
- Designate a Chief Information Security Officer (CISO)
- Implement multi-factor authentication (MFA) unless the CISO has approved other equivalent methods
- Encrypt nonpublic information both in transit and at rest unless the CISO has approved alternative controls
- Develop a written incident response plan
- Notify the DFS of security events such as data breaches within 72 hours of detection
- Submit to DFS an annual certification of compliance with the requirements of the Cybersecurity Regulation

For more information regarding the requirements under NYDFS Cybersecurity Regulations, see Part 500 Cybersecurity Regulations.

State Regulatory Guidance

State regulators have also weighed in on Big Data in financial services. For example, the New York Department of Financial Services issued Circular Letter No. 1, which provides guidance for using "external data sources" in underwriting life insurance. The letter highlights two main concerns around the use of algorithms and predictive modeling:

- That they will lead to unlawful discrimination in the affordability and availability of life insurance for protected classes
- That they will greatly reduce transparency in decision-making for consumers.

According to the letter, insurance providers must ensure that any underlying data source was not gathered discriminatorily and that the reasons behind any decisions are made available to the public. Notably, the guidance expressly states that insurers cannot use the proprietary nature of third-party algorithms to justify failing to explain unfavorable insurance decisions.

Healthcare

Below are federal laws in the healthcare sector that intersect with Big Data applications.

Health Insurance Portability and Accountability Act

There are two rules promulgated under the HIPAA that are particularly relevant to Big Data applications:

- The Privacy Rule (45 C.F.R. §§ 160, 164.500–164.534)
- The Security Rule (45 C.F.R. §§ 160, 164.302–164.318)

Under the Privacy Rule, protected health information (PHI) cannot be used or disclosed unless permitted by the rules or specifically authorized by the individual. There also are specific rules related to the sale of PHI or the use of PHI for marketing. The Security Rule sets forth detailed requirements for the protection of electronic PHI.

Practitioners should be aware of the fact that HIPAA is not a general medical privacy law. It does not apply to all healthcare data; instead, it applies to covered entities and their business associates. Covered entities are defined in the law as healthcare providers, health plans, and healthcare clearinghouses. Business associates are entities that provide services to covered entities where the performance of those services involves the use or disclosure of patient information.

Note, however, even though some healthcare data might not be regulated by HIPAA (as a result of not being processed by a covered entity or a business associate), that does not mean that it is not regulated. The comprehensive state laws discussed earlier in this practice note (like the CCPA) contain exemptions for data regulated under federal laws such as HIPAA, but, to the extent that information falls outside the scope of HIPAA, it may be regulated under those state laws. This is increasingly becoming an issue for health tracking technologies that leverage Big Data, such as smart watches and fertility apps.

HIPAA is enforced by the U.S. Department of Health and Human Services (HHS). For more general information regarding HIPAA, see HIPAA Privacy, Security, Breach Notification, and Other Administrative Simplification Rules.

State Healthcare Privacy Laws

There are a number of state laws that govern the privacy of health information and that may be relevant to a Big Data application that involves health information.

For example, the California Confidentiality of Medical Information Act (CMIA) builds upon the protections under HIPAA for California residents. CMIA requires that health records be created, maintained, and destroyed in a manner that preserves patient confidentiality. The law is enforced through both administrative fines and private suits, with nominal damages of $1,000 available even if actual damages cannot be shown.

Laws That Regulate Specific Technologies

Below are laws that regulate specific technologies that intersect with Big Data applications.

Biometric State Laws

Three states (Illinois, Texas, and Washington) have passed laws specifically regulating biometric information. The definitions of biometric identifiers in these laws vary but generally include data elements such as retina scans, iris scans, fingerprints, voiceprints, and facial geometry.

These laws require businesses to provide notice and obtain consent prior to collecting and sharing biometric information and also require businesses to implement data retention policies in relation to biometric identifiers.

Of the three states with biometric laws, Illinois's Biometric Information Privacy Act has the most requirements and is heavily litigated because it has a private right of action, along with statutory damages, and allows for attorney's fees.

In addition to these specific biometric laws, biometric information is increasingly regulated by U.S. privacy laws in other ways. The comprehensive privacy laws referenced earlier in this section (such as the CCPA) all regulate biometric information in some form. Additionally, state data breach notice laws are expanding to include biometric information in their definition of personal information. Even the FTC is paying increased attention in this area (see, e.g., the Everalbum settlement discussed above).

Employment Law and AI

New York and Illinois have passed laws that specifically regulate the use of AI in the hiring process. New York Local Law 144, which takes effect in April 2023, regulates the use of any automated tools in hiring and promotion decisions. The law requires employers and employment agencies to provide notice, an explanation of the source data used to build the tool, as well as the data retention policy, prior to the use of the tool. Covered entities must offer data subjects an opportunity to opt-out and request an alternative accommodation, as well as conduct an annual bias audit.

More narrow in application, the Illinois Artificial Intelligence Video Interview Act, which took effect in January 2020, applies to all employers that use AI specifically to process video interviews of Illinois applicants. It requires such employers to disclose the use of AI in the hiring process, provide each applicant an explanation of how AI is used, and obtain consent for use of AI, prior to the requesting video interview submissions. Further, it incorporates purpose limitation for the sharing and retaining of video data.

While the Illinois Act does not outline enforcement measures, in its current iteration a violation of the New York Law would result in penalties of up to $500 for first time violations and $1,500 per subsequent violations.

IoT Laws

As discussed earlier, Big Data and IoT are interrelated, and, as a result, IoT regulations impact how businesses collect Big Data. After largely being unregulated (or indirectly regulated), lawmakers are paying more attention to the issues posed by IoT devices in recent years, particularly as it pertains to cybersecurity concerns.

For example, congress recently passed The Internet of Things Cybersecurity Improvement Act of 2020. This law focuses on government procurement of IoT devices. It requires that the National Institute of Standards and Technology develop and publish guidelines on security standards for IoT devices used by government agencies. The Office of Management and Budget is responsible for reviewing the information security policies of each agency to ensure that they are in compliance with these standards. The standard could have broader application for industry if entities offering IoT devices begin to see the standard as a new baseline requirement.

At the state level, California and Oregon have passed laws regulating IoT devices. These laws cover not only devices that are sold to the government, but those generally sold within the state. They require IoT devices to implement "reasonable" security features to ensure the safety of such devices. It is possible that more states pass similar requirements in the near future, especially given that other states have historically tended to follow California's lead on privacy and security issues.

Federal Laws Relating to Children's Privacy

Below are federal laws relating to children's privacy that intersect with Big Data applications.

Children's Online Privacy Protection Act of 1998

The COPPA imposes rules and restrictions on operators of websites or mobile applications that collect personal

information online from children under the age of 13. To comply with COPPA, entities that collect personal information from children must (among other things) (1) provide consumers with a clear and prominent link to the company's applicable privacy policy, and (2) obtain (with limited exceptions) prior parental consent with any such collection. For more information on COPPA, see Children's Online Privacy Protection Act (COPPA) Compliance.

Family Education Rights and Privacy Act

The Family Education Rights and Privacy Act (FERPA) protects the privacy of education records. Family Education Rights and Privacy Act (FERPA), U.S. Department of Education. FERPA gives privacy rights related to education records to the parents of minors and then transfers these rights to the students once they turn 18. Parents or students have the right to review education records and correct records that are inaccurate or misleading. FERPA limits the individuals or organizations that educational records can be disclosed to without student or parent consent, except for "directory" information that can be disclosed publicly with only notice and not consent.

Other Applicable State Privacy Laws

Below are other applicable state privacy laws that intersect with Big Data applications.

Privacy Policy Laws

There is no federal law that requires the implementation of a privacy policy, but laws put in place by the states have made the policies broadly necessary and regulators like the FTC expect them. Any organization that collects or uses personal information for Big Data analytics should have a publicly available privacy policy that explains what information is collected, how it is used, and how it is shared. For example, the California, Virginia, and Colorado privacy laws all require privacy policies. In addition, the California Online Privacy Protection Act (CalOPPA) requires a clearly visible and accessible privacy policy and is generally seen as applying to all websites because of the likelihood that a public website will receive traffic from California residents. CalOPPA is enforced through California's Unfair Competition Law. Delaware state law also requires that websites, applications, and cloud computing services that collect personal information make their privacy policies conspicuously available. Del. Code tit. 6, § 205C.

Breach Notice Laws

All 50 states and the U.S. territories have laws that require private or government organizations to notify individuals and (frequently) state attorneys general of data security breaches that impact their personal information. Security Breach Notification Laws, National Conference of State Legislatures (Apr. 15, 2021). These laws specify who must comply with the law, the scope of personal information covered under the law, what qualifies as a data breach, notice requirements, and exceptions to the law. While these laws now exist in all U.S. jurisdictions, their specific provisions differ and will need to be analyzed individually by practitioners. Data Breach Notification in the United States and Territories, Privacy Rights Clearinghouse.

Litigation

Allegations of misuse of huge collections of data can generate numerous plaintiffs with many distinct claims. Perhaps the highest-profile Big Data litigation in progress is a multidistrict litigation (MDL) against Facebook for alleged misuse of data stemming from the Cambridge Analytica scandal. Consolidated in the Northern District of California, the MDL involves allegations that Facebook improperly shared users' personal information with third-party application developers and other business partners. In 2019, plaintiffs' key claims in a 414-page complaint survived a motion to dismiss. In re Facebook, Inc., 402 F. Supp. 3d 767, 776 (N.D. Cal. 2019).

Other litigation illustrates how the use of Big Data can implicate legal obligations that have nothing to do with data security or privacy. In a recent unpublished opinion, the Superior Court of Delaware ruled that GEICO violated state law by automatically evaluating insurance claims using rules based on a database of past claims. Green v. GEICO, 2021 Del. Super. LEXIS 308 (Sup. Ct. Del. Mar. 24, 2021). State law required a "reasonable investigation" of insurance claims "based on all available data." GEICO allegedly used rules to cap medical payments at the 80th percentile of claims in their database. And a claim would be denied entirely if GEICO's software determined it was for "passive" treatment more than eight weeks post-accident. The state court held that such mechanical decision-making based on historical data "did not process all available information and actually made investigations less likely to include all available information." The court made clear that updated automatic rules could be valid if they accounted for the proper factors, but that "human judgment should not be eliminated from the process" until a sufficiently nuanced system was in place.

Issues Related to De-identification

De-identification is the process of removing personal identifying information from a data set. De-identification Guidelines for Structured Data, Information & Privacy Commissioner of Ontario 1 (2016). Once data has been de-identified, it is no longer at risk of violating individual privacy and can be used for learning and research purposes. De-

identification provides a valuable balance between privacy and data use because it removes identifiers while retaining individual characteristics of the data set that can be helpful for research or analysis. The Value of De-identified Personal Data, Carnegie Mellon University, 2007. De-identification is often done using a risk-based approach that relies on the calculation of an acceptable level of risk of re-identification. A number of factors must be considered when determining that risk, including the audience the data will be released to, the types of identifiers in the original data, the likelihood that someone will try to re-identify the data, and the potential impact of disclosing general attributes of the entire group.

Standards for De-identified Data

HIPAA. Privacy and de-identification has become a growing interest across the biomedical and life sciences community. Raphael Chevrier et al., Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review, 21 J. Med. Internet Res. (2019). De-identification of health data mitigates privacy risks to individuals and allows for the use of Big Data for research and other secondary purposes. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, Health & Human Services (Nov. 12, 2012). Section 164.514 of the HIPAA Privacy Rule creates the standard for de-identification of PHI. 45 C.F.R. § 164.514. Two methods can be used to satisfy the de-identification requirements of HIPAA. The first option is a review of the de-identified data and determination of sufficient de-identification by a qualified professional. This is referred to as the Expert Determination standard. This standard requires the implementation of statistical or scientific principles and requires a minimal risk that the intended recipient could identify the individual. The second option is to guarantee removal of specific identifiers, when there is no actual knowledge that the remaining information could potentially be used to identify an individual. This de-identification approach is referred to as the Safe Harbor standard. There are 18 specific identifiers that must be removed if this is the chosen approach. HHS guidance provides further details on fulfilling these requirements, such as guidance on the necessary qualifications for an expert and the acceptable level of identification risk.

CCPA/CPRA. Personal information regulated under the CCPA/CPRA does not include information that has been de-identified or aggregated. Cal. Civ. Code § 1798.140(v)(3) (as amended). The CCPA/CPRA defines de-identified data as that which "cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer[.]"Cal. Civ. Code § 1798.140(m). In order to qualify for this exception from regulation, businesses that use de-identified data must also

- Take reasonable measures to ensure that the information cannot be associated with a consumer or household
- Publicly commit to maintain and use the information in deidentified form –and–
- Not to attempt to reidentify the information, except for the purpose of determining whether its de-identification process satisfies the CCPA/CPRA

Cal. Civ. Code § 1798.140(m).

Historically, the CCPA definition of de-identified data was challenging for some organizations to comply with, because the law includes no metrics by which to evaluate whether data can be used to "reasonably identify" individuals. However, as of January 1, 2021, the CCPA has been updated through Assembly Bill 713 (AB 713) to exclude data that has been de-identified in compliance with HIPAA standards. California Assembly Bill 713. While this update provided some clarity for healthcare organizations on the application of the CCPA to their data, it has also imposed some further burdens. The updated law now requires businesses to disclose whether they are selling HIPAA de-identified data and, if so, which methodology was used to de-identify the data.

AB 713 also clarifies that re-identified data must be regulated under HIPAA and other health privacy laws. The party that partakes in the re-identification will shoulder the burden of ensuring compliance with federal and state privacy laws. AB 713 further imposes requirements on contracts for the sale of de-identified data. These contracts must now include a prohibition on the re-identification of data if a party in the contract resides or does business in California.

Preventing Re-identification

As discussed above, the CCPA/CPRA includes several measures that attempt to prevent the re-identification of data. These include contractual prohibitions against re-identification and the imposition of regulatory liability on organizations that re-identify data. The CCPA/CPRA's requirement for procedures that prevent inadvertent disclosure of de-identified data could also be an effort to mitigate the risk of re-identification. The HIPAA Expert Determination standard also considers the risk of re-identification and bases the requirements for de-identification on an acceptable level of risk of re-identification.

Re-identification is a significant risk in instances where outside data sources can be used to de-identify a dataset. For example, a small set of purchase transactions could be used to re-identify individuals in a large, de-identified data set of transactions. The potential for re-identification is more dependent on what data outside organizations possess than on the capabilities of the original holder of the de-identified data.

Re-identification is typically achieved either when the original de-identification was insufficient, when pseudonyms used in de-identification are reversed, or when datasets are combined to reveal identities. Boris Lubarsky, Re-identification of "Anonymized" Data, 1 Geo. L. Tech. Rev. 202 (2017). Prevention techniques can be used to combat each of these three re-identification methods. The standards for re-identification in HIPAA and the CCPA/CPRA help ensure that de-identification will be sufficient. It is especially important to ensure sufficient de-identification of structured data sets, as indirect identifiers in these sets can lead to re-identification of an entire data set. If pseudonyms are used in de-identification, it should be ensured that they cannot be reversed. Use of a key and use of the same pseudonym for one individual in multiple instances can make it easier to re-identify data through pseudonym reversal.

When a de-identified data set is released, its anonymization can never be strengthened, only weakened. Boris Lubarsky, Re-identification of "Anonymized" Data, 1 Geo. L. Tech. Rev. 202 (2017). To protect privacy of personal data in large data sets, it is imperative that de-identification is performed to a high standard.

Big Data and Data Brokers

Data brokers are businesses that collect consumers' personal information and resell it to third parties. The data they provide about consumers can either form the basis of data sets used for Big Data analytics or augment existing data sets. Practitioners should be aware of how data brokers operate as well as the relevant laws and guidelines that apply to data brokers.

FTC Data Broker Report

The seminal report on data brokers was released by the FTC in May 2014. Titled Data Brokers – A Call for Transparency and Accountability, the report explains what data brokers are as well as how they use Big Data analytics in their work. Data brokers not only use raw data collected from various sources, but also can infer particular derived data from such information using highly specialized algorithms, such as a person's interests, orientations, hobbies, and spending habits. In the report, the FTC concludes that there exists a fundamental lack of transparency in data broker practices, as the bulk of their activity takes place without consumers' knowledge or informed consent. Given the above findings, the FTC in the report urges congress to improve data transparency through legislation that would empower consumers. Recommendations made by the FTC included enacting legislation that would allow consumers to learn about data broker practices, require data brokers to provide consumers access to the information held about them, require data brokers to provide notice of data collection to consumers, and require data brokers to disclose the sources of their data so that misinformation can be corrected.

The report also promulgated a set of best practices for data brokers. The FTC recommended that data brokers implement privacy-by-design, which includes taking privacy issues into account at every stage of product development. It also advised data brokers to implement effective measures to prevent the collection of personal information from children and teens, particularly when marketing goods and services. Lastly, data brokers should take reasonable precautions to ensure that downstream users do not use the data for discriminatory or other fraudulent purposes.

Once again, transparency and disclosure are key elements of compliance with FTC guidance. Counsel should carefully review their data broker clients' collection and dissemination practices, including those of the data brokers' clients who commercially exploit the information, to establish an effective strategy to protect consumer privacy.

Data Broker Laws

California and Vermont both have laws that regulate data brokers. See State Laws Related to Digital Privacy.

The California law defines data brokers as "a business that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship." Cal. Civ. Code § 1798.99.80. Both of these state laws require data brokers to register with the state and to make information associated with their registration, such as addresses and other business information, publicly available. In California, data brokers that fail to fulfill these requirements can be held liable through civil penalties, fees, and civil enforcement brought by the state Attorney General. The Vermont data broker law also requires that organizations publish a statement specifying details like the types of data collection and activities that a user may not opt-out from, and a statement about whether the data broker uses a purchaser credentialing process. Data brokers must also implement an information security program that protects personally identifiable information through administrative, technical, and physical safeguards.

In March 2020, the Vermont Attorney General brought the first data enforcement action under the state's data broker law. Divonne Smoyer, Samuel F. Cullari, & Alexis Cocco, Vermont Attorney General Brings First Data Broker Enforcement Action, Technology Law Dispatch (Mar. 17, 2020). Clearview AI was accused of amassing a database of billions of photographs and using facial recognition technology to create an identification service. The company purportedly violated the data broker law by using screen scraping technology to fraudulently acquire brokered personal information.

Mitigating Legal and Reputational Risks

Practitioners should consider the following questions when evaluating the risks associated with a matter that involves the use of Big Data analytics.

Data Format and Type

- Does the data contain transaction-level information or similarly granular data, or is the data aggregated in some manner?
- If the data contains transaction-level information or similarly granular data, is such data pseudonymized?
- If the data is aggregated or de-identified, have steps been taken to ensure that it cannot be re-identified?
- Does the data set include any personal information? If yes, what kinds of personal information are included in the data set?
- Does the data set contain sensitive personal information, for example, health information, financial information, or precise location?
- Does the data set include demographic information, such as gender, age, or race?
- Does the data set include information collected from children under 13? Children under 16?
- Does the data set include any information that relates to creditworthiness, credit standing, or credit capacity, such as defaults, income, credit scores, etc.?
- Does the data set include any other information that (1) bears on a consumer's character, general reputation, personal characteristics, or mode of living; and (2) was collected or used (even in part) for eligibility purposes for credit, insurance, or other commercial offerings?

Data Collection

- How was the data set obtained? Is it all first-party data, or does some of it come from third-party sources?
- Does the entity compiling the data and doing the analytics have a direct relationship with the consumer?
- Were APIs or web scraping used to obtain any of the information?

Contractual, Privacy Notice, and Other Restrictions

- Is the data use consistent with material promises made to consumers?
- Were the consumers provided material information about relevant data practices?
- Are reasonable measures being undertaken to know the purposes for which customers are using the data? For example, have you taken reasonable precautions to ensure that downstream users do not use the Big Data products for discriminatory or other fraudulent purposes?
- If a third party is providing the data, does the agreement with that third party place any restrictions on the use of the data?

Other Questions

- If you compile Big Data for others who will use it for eligibility decisions (such as credit, employment, insurance, housing, government benefits, and the like), are you complying with the accuracy and privacy provisions of the FCRA?
- If you receive Big Data products from another entity that you will use for eligibility decisions, are you complying with the provisions applicable to users of consumer reports?
- If you are a creditor using Big Data analytics in a credit transaction, are you complying with the requirement to provide statements of specific reasons for adverse action under ECOA?
- Are you complying with ECOA requirements related to requests for information and record retention?
- If you use Big Data analytics in a way that might adversely affect people in their ability to obtain credit, housing, or employment (1) are you treating people differently based on a prohibited basis, such as race or national origin?; or (2) do your policies, practices, or decisions have an adverse effect or impact on a member of a protected class, and if they do, are they justified by a legitimate business need that cannot reasonably be achieved by means that are less disparate in their impact?
- Do you have reasonable safeguards in place to protect consumer information that are appropriate for the amount and sensitivity of the data at issue, the size and complexity of the company's operations, and the cost of available security measures?
- Have you reviewed your data sets and algorithms to ensure that hidden biases are not having an unintended impact on certain populations?
- Have you confirmed the accuracy of your predictions based on Big Data?
- Have you considered whether fairness and ethical considerations advise against using Big Data in certain circumstances?

Future of Big Data

Practitioners working in Big Data should make sure to stay on top of new developments in privacy law that might affect their matters. For example, there are now three states with comprehensive privacy laws, and it is likely that there will be more in the near future. Practitioners will need to understand whether and how these laws apply to Big Data and be aware of how to reconcile different legal requirements across various legal regimes. In addition, there is the potential for a comprehensive federal privacy law to pass that could change the landscape for Big Data, or the FTC could use its authority to promulgate rules that touch on Big Data concerns. Finally, practitioners should stay on top of enforcement actions brought by the FTC and other regulators involving Big Data as those will provide insight into how regulators are approaching Big Data issues.

Kirk Nahra, Partner, Wilmer Cutler Pickering Hale and Dorr LLP

Kirk Nahra has been a leading authority on privacy and cybersecurity matters for more than two decades. Indeed, he is one of the few lawyers in the world ranked in Band 1 by *Chambers* in privacy and data security. He is also the winner of the 2021 Vanguard Award from the International Association of Privacy Professionals (IAPP)—one of the most prestigious in the privacy field—which recognizes one IAPP member each year who demonstrates exceptional leadership, knowledge and creativity in privacy and data protection. Mr. Nahra counsels clients across industries, from Fortune 500 companies to startups, on implementing the requirements of privacy and data security laws across the country and internationally, and he advocates for clients experiencing privacy and security breaches. Mr. Nahra also represents clients in contract and deal matters, enforcement actions, litigation and investigations related to a wide range of issues before the Federal Trade Commission (FTC), the US Department of Health and Human Services (HHS) Office for Civil Rights, and other state and federal privacy and security regulators.

Mr. Nahra is best known for his work with health insurers, hospitals, service providers, pharmaceutical manufacturers and other health care industry participants. He has a deep understanding of the privacy and security issues healthcare companies face relating to HIPAA rules, state and federal legislation, enforcement activities, internal investigations, international principles, due diligence in transactions, data breach risk assessments, and the key lines between regulated and unregulated data. During his decades of experience, Mr. Nahra has developed compliance programs, drafted privacy and information security policies, negotiated agreements involving health data, responded to health incidents and defended clients against government investigations.

In recent years, Mr. Nahra has represented technology companies, advertising service providers, financial services companies, hospital systems, health insurers, healthcare technology companies, consumer products companies and others in front of the FTC, the HHS Office for Civil Rights, and other privacy and security regulatory agencies. He advises clients on avoiding privacy and security investigations, navigating situations where investigations are likely, and then handling both the actual investigation and related issues involving consumers, customers, legislators, regulators and others.

Mr. Nahra also has substantial experience working with clients in the financial services and insurance industries on privacy and data security matters relating to the Gramm-Leach-Bliley Act, Fair Credit Reporting Act, Fair and Accurate Credit Transactions Act, data aggregation and sharing practices, and privacy and data security compliance under a wide range of state and federal laws. He also has a breadth of experience drafting and evaluating data security practices and policies across varying industry standards, has investigated and litigated potential fraud against insurers, and has assisted with the development and oversight of corporate compliance programs.

Additionally, Mr. Nahra is well versed in a variety of other privacy and consumer protection issues, including marketing laws pertaining to email, phone and online communications; the Children's Online Privacy Protection Act; and the Family Educational Rights and Privacy Act of 1974.

Professional Activities

A leader in the privacy bar, Mr. Nahra has been involved in developing the privacy legal field for 20 years. As a founding member and longtime board member of the International Association of Privacy Professionals, he helped establish the organization's Privacy Bar Section and their first and most popular certification for Certified Information Privacy Professionals. He is a member of the Center for Cybersecurity and Privacy Protection National Advisory Board. He has taught privacy issues at several law schools, including serving as an adjunct professor at the Washington College of Law at American University and at Case Western Reserve University. In addition, he currently serves as a fellow with the Cordell Institute for Policy in Medicine & Law at Washington University in St. Louis and as a fellow with the Institute for Critical Infrastructure Technology. He actively shares his privacy insights through numerous speeches and articles, and on social media.

Arianna Evers, Special Counsel, Wilmer Cutler Pickering Hale and Dorr LLP

Arianna Evers advises and advocates for clients on privacy, data security and consumer protection issues arising under federal, state and international laws.

Ms. Evers represents clients in investigations and litigation with state attorneys general concerning alleged privacy and consumer protection violations, and in enforcement actions and regulatory investigations brought by the Federal Trade Commission under its Section 5 Authority. Ms. Evers advises clients on their obligations under federal and state data breach notification laws and coordinates data breach investigations, including working with forensic firms and providing notice to regulators and affected individuals. She also consults on privacy and data security issues for congressional inquiries, including preparing senior executives for hearings and meetings with Capitol Hill staff.

Ms. Evers leverages her many years as a trial attorney for technology companies in her work advising clients on how to identify and mitigate regulatory and contractual risks. Ms. Evers regularly counsels clients on their compliance obligations and risks relating to federal laws, including the Gramm-Leach-Bliley Act, Health Insurance Portability and Accountability Act, CAN-SPAM, Children's Online Privacy Protection Act, and Fair Credit Reporting Act; state laws such as the Massachusetts Data Security Regulation, NYDFS cybersecurity regulations and the California Online Privacy Protection Act; and international laws like the EU General Data Protection Regulation. She also advises companies on legal, contractual and reputational risks relating to data mining and other practices involving big data.

Ms. Evers' practice also includes a broad range of transactional work, including negotiating privacy and data security provisions in agreements, drafting privacy policies and other consumer-facing disclosures, and conducting due diligence for corporate acquisitions and mergers.

Ms. Evers began her career as a member of the firm's litigation group, where she focused on intellectual property and other complex commercial litigation and counseling at the trial and appellate stages. Ms. Evers has extensive experience in all phases of complex litigation and has shepherded multiple cases from complaint through trial and post-trial briefing. Ms. Evers has argued in court, prepared expert witnesses for trial, prepared direct- and cross-examinations, taken and defended depositions, and handled all aspects of pretrial discovery and motion practice.

Ms. Evers has experience serving clients in wide range of industries, including financial services, Internet-related services and products, marketing, insurance, automotive, computer software, telecommunications, semiconductor, healthcare, pharmaceutical, biotechnology, and many other sectors.

Past Experience

Prior to starting at WilmerHale, Ms. Evers volunteered as a staff attorney with the American Civil Liberties Union of Massachusetts, where she worked on a variety of civil rights matters, with a particular focus on First Amendment controversies. Ms. Evers also served as the communications director on a successful reelection campaign for Boston City Councilor At-Large.

Prior to law school, Ms. Evers counseled clients on crisis management at a large public affairs firm in Washington DC. She also served on the national advance staff of a US presidential campaign.

Professional Activities

Ms. Evers is admitted to practice before the US Court of Appeals for the Federal Circuit and the First Circuit, and the US District Court for the District of Massachusetts.

Ali Jessani, Senior Associate, Wilmer Cutler Pickering Hale and Dorr LLP

Ali A. Jessani counsels clients on the privacy, cybersecurity and regulatory risks presented by new and proposed uses of technology and consumer information. Specifically, he advises clients with compliance issues related to the California Consumer Privacy Act, the General Data Protection Regulation, the Gramm-Leach-Bliley Act, the Fair Credit Reporting Act, state biometric laws and other federal and state laws governing data sharing, ownership and protection. Mr. Jessani also guides companies through legal obligations after data breaches, as well as through state and federal regulatory investigations.

While pursuing his legal education, Mr. Jessani was an extern in the US Department of Justice Civil Rights Division's Voting Rights Section and an intern in the Voter Expansion Department of the Democratic National Committee. He was also Executive Editor of the *Duke Journal of Gender Law and Policy*.

Ali Jessani, Senior Associate, Wilmer Cutler Pickering Hale and Dorr LLP

Ali A. Jessani counsels clients on the privacy, cybersecurity and regulatory risks presented by new and proposed uses of technology and consumer information. Specifically, he advises clients with compliance issues related to the California Consumer Privacy Act, the General Data Protection Regulation, the Gramm-Leach-Bliley Act, the Fair Credit Reporting Act, state biometric laws and other federal and state laws governing data sharing, ownership and protection. Mr. Jessani also guides companies through legal obligations after data breaches, as well as through state and federal regulatory investigations.

While pursuing his legal education, Mr. Jessani was an extern in the US Department of Justice Civil Rights Division's Voting Rights Section and an intern in the Voter Expansion Department of the Democratic National Committee. He was also Executive Editor of the Duke Journal of Gender Law and Policy.