

Chapter 66

Big Data: Legal and Compliance Considerations for Investment Managers

Nathan J. Greene*

Partner, Sidley Austin LLP

Colleen Theresa Brown

Partner, Sidley Austin LLP

[Chapter 66 is current as of August 23, 2020.]

- § 66:1 Overview
- § 66:2 Introduction to Data, Alternative Data, Big Data, and the Related Disciplines of Artificial Intelligence and Machine Learning
 - § 66:2.1 Data
 - § 66:2.2 Alternative Data
 - § 66:2.3 Big Data
 - § 66:2.4 Artificial Intelligence and Machine Learning
- § 66:3 Investment Management Functions Making Heavy Use of Data
 - § 66:3.1 Investment Analysis
 - § 66:3.2 Trading
 - § 66:3.3 Marketing
 - § 66:3.4 Robo-Advice

* Portions of this chapter were previously published as articles by Mr. Greene. *See, e.g., Nathan J. Greene, Big Data: A Legal and Compliance Guide for Investment Managers, THE INV. LAW.* (June 2019).

- § 66:3.5 RegTech
- § 66:3.6 Operations and Reporting
- § 66:4 New Sources of Data Used by Investment Management Firms
 - § 66:4.1 Internet Content
 - § 66:4.2 Social Media
 - § 66:4.3 Online “Exhaust”
 - § 66:4.4 Internet of Things
 - § 66:4.5 Geospatial/Drones
 - § 66:4.6 Consumer Transactions
 - § 66:4.7 Location
 - § 66:4.8 Biometric
- § 66:5 U.S. Legal Principles and Regulatory Views and Trends
 - § 66:5.1 Data Protection and Cybersecurity
 - [A] Avoiding Loss
 - [B] Protection Against Infringement
 - § 66:5.2 Legally Protected Data—Privacy
 - [A] GDPR
 - [B] CCPA
 - [C] A Note on Deidentified or Aggregated Data
 - § 66:5.3 Select Additional Examples of Protected Data—Government Data, National Security Data, and the CFAA
 - [A] Government Data
 - [B] National Security
 - [C] Web Scraping and the Computer Fraud and Abuse Act (CFAA)
 - § 66:5.4 Insider Trading
 - § 66:5.5 Artificial Intelligence Techniques
 - [A] U.S. Treasury Report
 - [B] FINRA Guidance
 - [C] SEC Guidance
 - [D] U.S. Federal Reserve Guidance
 - [E] Source Code Access
 - § 66:5.6 Antitrust
 - § 66:5.7 Bias
 - § 66:5.8 New York’s Martin Act and General “Fairness” Principles
- § 66:6 Data-Related Compliance and Controls Considerations
 - § 66:6.1 Data Sourcing
 - [A] Vendor Diligence
 - [B] Data Lineage
 - [C] Personal Information
 - [D] Web Scraping/CFAA
 - § 66:6.2 Selling Data
 - § 66:6.3 Data Governance
 - § 66:6.4 Model Governance
 - § 66:6.5 Robo-Advice Considerations
 - § 66:6.6 Diligence of RegTech Providers
 - § 66:6.7 Conflict of Interest Considerations and Disclosure—Whose Data Is It?

§ 66:1 Overview

Every industry is expanding its use of data, tapping new information sources, and applying ever more sophisticated data analytics to develop and monetize new sources of insight. It should be no surprise that investment managers, along with other professional investors and traders, are at the forefront of these trends.

Faced with an expanding range of novel information sources and increasingly varied applications of data—including artificial intelligence (AI) and similar computing methods powered by data—it is critical that investment management lawyers and compliance professionals have both a solid working understanding of their firms' data-related activities, the attendant legal and operational risks, and a framework of thinking to apply to them.

An opening question that many legal and compliance professionals will have is whether—and if so, how—the use of big data, alternative data, and the related disciplines of machine learning and AI are regulated activities for investment managers. To begin the conversation, consider these observations on machine learning by Robert Pozen (Mr. Pozen is a lawyer by training and formerly a top executive at two prominent mutual fund firms):

ML [machine learning, a variant of artificial intelligence] is particularly adaptable to securities investing because the insights it garners can be acted on quickly and efficiently. By contrast, when ML generates new insights in other sectors, firms must overcome substantial constraints before putting those insights into action. For example, when Google develops a self-driving car powered by ML, it must gain approval from an array of stakeholders before that car can hit the road. These stakeholders include federal regulators, auto insurers, and local governments where these self-driving cars would operate. *Portfolio managers do not need regulatory approval to translate ML insights into investment decisions. (emphasis added)*¹

Pozen's closing remark, that "portfolio managers do not need regulatory approval" to proceed, requires scrutiny. As a literal matter, he is right, at least in the case of U.S. regulation of investment managers. No direct approval is required to deploy data and AI in an investment program. But a more complete assessment will quickly recognize that many different legal principles do, in fact, constrain and regulate this area. This chapter is intended to help the legal and compliance

1. Robert Pozen & Jonathan Ruane, *What Machine Learning Will Mean for Asset Managers*, HARV. BUS. REV. (Dec. 3, 2019), <https://hbr.org/2019/12/what-machine-learning-will-mean-for-asset-managers>.

practitioner see the shape of the patchwork. To that end, the chapter begins with an introduction to the concepts of data, alternative data, big data, and AI. The discussion then covers examples of an organization's data users, likely sources of data, and organizational controls for data collection and processing. A review of the ways different types of data are regulated is also provided.

§ 66:2 Introduction to Data, Alternative Data, Big Data, and the Related Disciplines of Artificial Intelligence and Machine Learning

This section introduces basic technical elements to the concepts of data, alternative data, big data, and AI. A key take-away should be that data frequently are available at such scale that their full value can be captured only with sophisticated quantitative techniques. Data and AI initiatives thus tend to develop in tandem, with related legal and compliance issues closely joined and overlapping. Legal and compliance professionals should anticipate analyzing these topics together.

§ 66:2.1 Data

Data as understood today exist largely as intangibles, a stream of digitized information. Yet data also are high-value assets to be used, bought, and sold.

Data can be gathered from myriad sources and exist in both their "raw state" and in various states of organization or disorganization (or as data professionals prefer, states of categorization or manipulation). Data can be presented with all links to the source intact or with various levels of deidentification, aggregation, or anonymization. New data can be created through the manipulation of data or as metadata or data markers (for example, when and how data were created). This "derived" or "resultant" data can be as useful, and important, as the original source data.

§ 66:2.2 Alternative Data

Types of data consumed by investment professionals today are broad and expanding daily. Examples include credit card spending; money transfer patterns; weather; traffic; port or other infrastructure activity; utility and cell phone usage; geolocation; online search statistics; news or social media "sentiment" analysis; and more.² Collectively, the investment management industry terms these "alternative data," with the name referring both to the novelty and variety

2. *Infra* section 66:3 reviews types of data being accessed by investment analysts today in more detail.

of the data and in specific contrast to more “traditional” data coming from sell-side broker-dealer and bank research desks.

§ 66:2.3 **Big Data**

Alternative data, in turn, often are examples of big data—“extremely large datasets that may be analyzed computationally to reveal patterns, trends and associations.”³ Stated differently, the meaning in big data is not plain; information is revealed only through analysis. Another common way to approach the concept is through what are widely referred to as the “five Vs” of big data:

- (1) Volume, referring to scale;
- (2) Variety, referring to many types of data;
- (3) Velocity, referring to speed of collection and processing;
- (4) Veracity, referring to quality (that the data should be reliable); and
- (5) Value, referring to the idea that each of the other “Vs” should build on each other to drive actionable, cost-efficient analysis and insight.

This is sometimes pictured as a pyramidal relationship in which volume and variety of data are the wide base of the structure, and value is the apex, derived from the contributions of each layer below it.

§ 66:2.4 **Artificial Intelligence and Machine Learning**

A technical discussion of AI is well beyond the scope of this chapter, but as a general matter, the phrase captures a spectrum of quantitative data analysis techniques with a common output: the machine is equipped with the ability to perform tasks historically understood to require natural intelligence (image identification, pattern recognition, categorization, language translation and analysis, strategy games, etc.).

Importantly, even the most fully realized AI in existence today is at most a *simulation* of human intelligence, that is, it *appears* intelligent. In some cases, the machine appears to “learn” as well. It recalibrates and gets progressively better at its task without all improvements being specifically programmed. Progressive, unprogrammed, or partially programmed improvement in the functions of an algorithm is what is referred to as machine learning or ML.

3. *Big data*, OXFORD ENGLISH DICTIONARY (2020), <https://www.oed.com>. The term was added to this venerable volume only in 2013. To help place the moment, other tech-focused words added that year include crowd-sourcing and e-reader.

But the machine is in all cases domain-limited. It operates under a fundamentally narrow set of constraints and limitations. AI-practitioners term this “weak or narrow AI.” By contrast, strong AI (or “deep or general AI”) would be a single application with the ability to learn and solve for a wide array of problems.⁴

The power and promise of AI have generated deep policy and commercial interest. The U.S. government, in particular, has for many years recognized the need to foster these new technologies with a healthy amount of care and scrutiny for the potential harms they may cause, in particular to ensure data integrity, security, and reliability, and control for bias and discriminatory application or impact. The Obama administration brought specific focus and energy to AI policy, including issuing a report focused on AI use for the public good in *Preparing for the Future of Artificial Intelligence* in October 2016, and a follow up report in December 2016, *Artificial Intelligence, Automation, and the Economy*, focused largely on private sector impacts.⁵ A key point of these reports is the incredible disruptive power of AI on businesses in every sector of the economy—financial services and investment management being no exception.

-
4. Many industry publications provide general overviews of artificial intelligence and treat AI as a spectrum of techniques in the manner described here. For additional treatment of the topic, see ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN ASSET MANAGEMENT, BLACKROCK VIEWPOINT (Oct. 2019) [hereinafter BlackRock AI Whitepaper], <https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-artificial-intelligence-machine-learning-asset-management-october-2019.pdf>; ARTIFICIAL INTELLIGENCE (AI) IN THE SECURITIES INDUSTRY, FINRA WHITEPAPER (June 10, 2020) [hereinafter FINRA AI 2020 Report], <https://www.finra.org/rules-guidance/key-topics/fintech/report/artificial-intelligence-in-the-securities-industry/key-challenges>.

Also of interest, the U.S. Commerce Department’s National Institute for Standards and Technology (NIST) has a program focused on the support and expansion of AI in a wide array of applications. For a description of NIST initiatives on AI, see NAT’L INST. FOR STANDARDS & TECH., MACHINE LEARNING/A.I., <https://www.nist.gov/project-category/materials-genome-initiative-mgi/machine-learning-ai>; and NAT’L INST. FOR STANDARDS & TECH., ARTIFICIAL INTELLIGENCE, <https://www.nist.gov/topics/artificial-intelligence>. NIST’s intergovernmental counterpart, the International Standards Organization (ISO), has a similar program, see Int’l Standards Org, ISO/IEC JTC 1/SC 42, ARTIFICIAL INTELLIGENCE, <https://www.iso.org/committee/6794475.html>.

5. EXEC. OFF. OF THE PRES., NAT’L SCI. & TECH. COUNCIL COMM. ON TECH., PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (Oct. 2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf; EXEC. OFF. OF THE PRES., ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY (Dec. 2016), <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>.

Investment in AI increasingly is seen as a national security imperative to strengthen national advantages in the global economy, with the United States, China, and the European Union in a race for dominance. In February 2020, the European Commission published a whitepaper on the use of AI as part of the European Union's digital strategy.⁶ The whitepaper urges a common approach to AI development to increase EU leadership and use of AI, with focus on development of a system of safeguards to foster trust in higher risk applications. The same month, President Donald Trump issued an *Executive Order on Maintaining American Leadership in Artificial Intelligence* on February 11, 2019.⁷ The executive order emphasizes the need to drive R&D on AI to maintain global and economic advantages, as well as to ensure trust in the technology and that AI development preserves American values. To pursue those goals, one of the key strategies of the executive order is to support enhanced access to high quality data.

§ 66:3 Investment Management Functions Making Heavy Use of Data

An investment management legal and compliance professional considering the implications of data for her organization will want to understand which parts of the organization are important data users. For many organizations, it should not be surprising to find that data are integral across the business. These examples can offer a starting point.

§ 66:3.1 Investment Analysis

An anecdote from a McKinsey report draws a picture of how alternative data is being used by analysts. The report describes a real-estate fund manager conducting traditional property- and market-level

-
6. European Comm. White Paper on Artificial Intelligence—European Approach to Excellence and Trust (Feb. 19, 2020), <https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-european-approach-excellence-and-trust>. In addition to its announced intention to build an EU-wide strategy to foster and manage adoption of AI, the European Commission anticipates an EU strategy to foster and manage a comprehensive data economy. See EUROPEAN COMM'N, EUROPEAN DATA STRATEGY: MAKING THE EU A ROLE MODEL FOR A SOCIETY EMPOWERED BY DATA, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.
 7. Executive Order on Maintaining American Leadership in Artificial Intelligence (Feb. 11, 2019), <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>.

analysis to develop rent forecasts for the firm. The manager then layers on “Yelp reviews, information on foot traffic, and credit-card spending data.”⁸ This may not seem revolutionary—the manager could have posted employees with clipboards and survey sheets outside buildings to develop similar information—but the point is that these data are more readily accessible and scalable, less prone to gaps and errors, and constantly refreshing; in other words, the same at a surface level, but categorically different.

Given the intuitive logic that new and better data create new value, legal and compliance professionals should assume that their firms’ investment analysts are using alternative data. The questions then will be how much, from how many different sources, and subject to what controls.⁹

§ 66:3.2 *Trading*

Automated or semi-automated trading come in two basic flavors, one in which trading and the investment process are tightly linked, with quantitative investment analysis feeding directly to market execution, and another in which trading is a standalone function, discrete and separate from research and analysis, but still fully optimized by sophisticated software applications. Either way, the legal and compliance professional assessing a firm’s data practices should recognize that the underlying trading programs will be grounded in an array of market data feeds interacting with trade execution analysis, making trading desks and trading software longstanding and prodigious consumers of data.

§ 66:3.3 *Marketing*

Data-driven approaches also have been transforming marketing. A common first step is harnessing data already in the system. For example, a client or prospect’s “profile” is comprised of metrics such as client type, size, location, channel, transaction and account information (past and present), and points of contact, etc. Historically, much of this was siloed, resulting in small or fragmented datasets unlikely to present a full picture and certainly unlikely to reveal predictive information. Increasingly, however, marketing software can cross-compare large pools of data to begin to identify signals for when a client is disposed to increase or reduce investment with the firm,

8. MCKINSEY & CO., *ADVANCED ANALYTICS IN ASSET MANAGEMENT: BEYOND THE BUZZ* (Mar. 20, 2019), <https://www.mckinsey.com/industries/financial-services/our-insights/advanced-analytics-in-asset-management-beyond-the-buzz>.

9. For a review of possible controls, see *infra* section 66:6.

contact the firm with questions (or expects to be contacted), or is otherwise ripe for fresh consideration by the client coverage team. Marketing teams also look outside the firm for data on clients and prospects. Social media is a prime example, and various tools gather social media data specifically for this purpose.

As will be suggested in section 66:5 below, these types of data-driven marketing initiatives often involve personal data (information on natural persons), which present some of the most sensitive data handling requirements and implicate a variety of state, federal, and international laws.

§ 66:3.4 **Robo-Advice**

The rise of model-based approaches to delivering customized investment advice to a wider audience at lower cost often is termed “robo-advice” (or sometimes “digital advice”). The gist of the service is that after the client completes a detailed online questionnaire, an algorithm should be able to provide a reasonably tailored investment program to the client without the expense of human judgment and handholding. It is perhaps self-evident, but a considerable volume of personal information on users must be collected, maintained, and protected in the course of operating a robo-advice platform.

Of course, these services come on a spectrum and rarely purport to entirely replace the role of human-led investment advice, as an industry whitepaper observes:

Digital advice can be used to supplement traditional advisors, and most digital advisors offer multiple ways to engage with a human professional. While some digital advice firms offer a greater degree of human supervision of client services and trading systems than others, digital advisors have a fundamental obligation to oversee their systems and mitigate risks associated with digital processes. Rather than replacing human advisors, digital advisors can help them automate processes and more effectively provide advice at scale.¹⁰

There also will be next generation robo-advice models that draw on new sources of data and pose new questions. Consider a service that mines social media or online search activity for greater insights into the client’s circumstances or risk tolerances. In one version of the service, it could cross-check learning from a client’s social media accounts against the questionnaire and highlight potential inconsistencies. In another version, the questionnaires themselves might be made “smart” and adapt seamlessly to the client, even asking different

10. BlackRock AI Whitepaper, *supra* note 4.

types of questions based on that social media learning (in the same way that different users of many online services can see quite different versions of the service tailored to the individual). In another version, the service would pitch additional products based on that learning (for example, suggesting college or health savings accounts, annuities, or other offerings based on apparently relevant personal information).

§ 66:3.5 *RegTech*

How data and AI inform a firm's control functions, especially around regulatory compliance, often is referred to as RegTech. The idea is simply that technology, especially when it can analyze data and surface anomalies and correlations more efficiently than human eyes and intuition, is part of today's compliance officer toolkit. These technologies support controls to identify fraudulent transactions, insider trading risks, trade restriction violations, cybersecurity vulnerabilities, and a wide array of other regulatory concerns.

A significant driver for investment in RegTech is the perception of an arms race. Regulators trumpet their success in developing quantitative and risk analytic processes that crunch industry data and guide their regulatory inspections, rulemaking, and other initiatives. Meanwhile, compliance officers and industry executives are riveted by the possibility that their regulators might "know their data better than they do." Accelerating implementation of RegTech then becomes necessary simply to keep up.¹¹

§ 66:3.6 *Operations and Reporting*

Increased regulation, often with the aim of monitoring both firm-specific risk and potential broad market risk, has significantly increased regulatory reporting across the investment management industry. Requirements added within the past decade just by the SEC include Form N-PORT, Form N-CEN, and Form PF (as well as new, more data-intensive sections of Form ADV). In the past, data aggregation, management, and similar processes could be something of a sideshow (or thought of as purely back office), but now often are integrated across front- and middle-office workflow processes.

11. The perceived data analytics arms race does not just involve the risk that regulators will know a firm's data better than the firm does. As increasing volumes of information about investment managers are required to be made public, and at increasing frequency, there is also risk that competitors, journalists, academic researchers, and others are drawing conclusions about a firm based on its public data. Forward thinking about those constituencies is part of the race.

As an illustration of scale, the industry whitepaper cited above describes one large firm's portfolio reporting and risk analytics system as creating and reviewing over one million daily risk and exposure reports, which in turn are rooted in the firm's massive internal datasets.¹² Here again, sophisticated data analysis techniques based on AI/ML models automate previously manual processes.

§ 66:4 New Sources of Data Used by Investment Management Firms

Thoughtfully assessing from where an organization's data are sourced is another important step for the legal and compliance professional. Ideally, each functional group will have a ready source catalog, and as covered in section 66:6 below, there will be organizational controls in place for how data are collected and processed.

Section 66:5 also reviews in more detail how these different types of data are regulated, with the most common source of regulation being personal privacy laws.

Likely sources of data are discussed in the following sections.¹³

§ 66:4.1 Internet Content

No information source can compete with the richness and depth of the internet, and new tools have made internet data accessible and consumable at scale. These include web scraping (the automated, systematic capture of internet pages by specialized software) and natural language processing or NLP techniques (which allow computers to read and analyze text).

The industry whitepaper cited above describes the power of combining web scraping and NLP techniques. The paper refers to an analyst team charged with contributing investment ideas for a strategy that tracks the Russell 3000 Index and observes that it is simple math that the team will consume roughly 12,000 quarterly and annual reports each year. While a mammoth undertaking by any measure, increasingly these reports are captured and classified on an

12. BlackRock AI Whitepaper, *supra* note 4.

13. At this point, none of the types of data highlighted here should be considered obscure or novel. A recent survey suggests that a majority of hedge fund firms today use at least one of the following in their investment processes: consumer transaction data, social media data, web-scraped data, app usage data, and internet of things data. The same survey found that 40% of firms use geolocation data and 20% use satellite imagery data. Charlie Marlow, *Survey Finds Widespread and Increasing Use of Alternative Data by Hedge Funds*, HEDGE FUND L. REP. (Oct. 17, 2019).

automated basis by web scraping and then analyzed in real time with sophisticated NLP capability.¹⁴

Web scraping, while common practice, is subject to considerable legal debate, most commonly around the application of the Computer Fraud and Abuse Act (CFAA), as discussed in sections 66:5 and 66:6 below.

§ 66:4.2 Social Media

Social media networks represent a critical subset of online activity. By their nature, these sites invite users to generate staggering volumes of data.¹⁵ Web scraping, NLP, and various tracking technologies, supplemented here by AI image recognition software, make these troves of information accessible.

As already suggested, social media data can be intrinsically personal, so potentially quite sensitive. Social media site terms of use and user privacy settings also present important restrictions to access by third parties.

§ 66:4.3 Online “Exhaust”

Online activity is itself a data source as users generate “data exhaust,” a term of art referring to data created as a byproduct of a person’s online actions (*for example*, sites visited, links clicked, time on a page, time of day records, mobile versus computer¹⁶). Cookies and other tracking technologies record these actions and make them available for subsequent analysis.

The California Consumer Privacy Act (CCPA), further discussed in section 66:5 below, can treat unique, persistent identifiers such as those contained in cookies as a type of “personal information” for which businesses must provide transparency and certain data rights. There has been controversy around when a site operator allowing a third party to place cookies on the site could constitute a “sale” of personal information within the meaning of the CCPA, and major web browsers have increased cookie controls and even default settings to block third-party cookies in recent years.¹⁷ Looking ahead, new privacy laws and regulations, technological trends, and user preferences

14. BlackRock AI Whitepaper, *supra* note 4.

15. To illustrate scale, 2019 estimates assume Facebook users consume 100 million hours of daily video watch time, upload 350 million photos daily, and generate 4 million likes every minute.

16. Search data is a variant of online exhaust. Again, to illustrate scale, Google in 2020 is estimated to generate three to six billion searches daily.

17. *See, e.g.,* Sara Morrison & Rani Molla, *Google’s Cookie Ban Is Good News for Google—And Maybe Your Privacy*, VOX (Jan. 16, 2020), <https://www.vox.com/recode/2020/1/16/21065641/google-chrome-cookie-ban-advertisers>; Sara Morrison, *Apple Is Finally Making It Easy to Hide*

for transparency and choice over cookies may affect the volume and quality of data available in the market.

§ 66:4.4 **Internet of Things**

The “internet of things” (IoT) refers to the interconnection of objects that were previously separate and static and thus not easily identified or monitored, but are now provided with connectivity through a private defined network or, frequently, the internet. “Smart” devices—televisions, printers, thermostats, appliances, industrial equipment, pipelines, etc.—are equipped with sensors that generate constant data feeds. Resulting data can be used for purposes connected to and supporting the physical device’s primary utility, such as inventory management, usage monitoring, software updates, maintenance, and the like, but also present opportunity for secondary use and are routinely monetized at a granular or aggregate level. These data can be of great interest to investment professionals.

§ 66:4.5 **Geospatial/Drones**

Satellite imagery has been widely used by investment management firms for years, and growing numbers of drones expand the scope of available images. Now, AI image recognition software has transformed the process, improving efficiency.

Drones present a variety of legal issues from trespassing laws, to Federal Aviation Administration (FAA) regulations, to privacy.¹⁸

§ 66:4.6 **Consumer Transactions**

A major source of data is from consumer (and business) transactions. Transaction data can be developed from many sources, including merchant systems, credit card companies and other processors, banks, and financial services websites and apps. These include personal finance aggregation services that bring together data on an individual’s credit-card and debit-card transactions, bank accounts, investment accounts, and loan accounts.

from *Trackers*, VOX (June 22, 2020), <https://www.vox.com/recode/2020/6/22/21299398/apple-ios14-big-sur-privacy-wwdc-2020>.

While this chapter focuses principally on U.S. law issues, cookies also have been subject to analysis in the European Union. See, e.g., UK INFO. COMM’RS OFF., GUIDANCE ON THE USE OF COOKIES AND SIMILAR TECHNOLOGIES (July 2019), <https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/>.

18. See Virginia K. Trunkes, *Balancing New Technology and Privacy When Using Drones in Land Use and Construction*, NAT’L L. REV. (June 26, 2020), <https://www.natlawreview.com/article/balancing-new-technology-and-privacy-when-using-drones-land-use-and-construction>.

Financial privacy laws regulate how financial institutions handle nonpublic consumer financial information and consumer profiles. At the federal level, examples include the Bank Secrecy Act, the Gramm-Leach-Bliley Act (including its Regulation S-P, to which most investment managers are subject), and the Fair Credit Reporting Act. State laws, including the California Financial Information Privacy Act, also protect personal financial information and likewise can restrict disclosure and repurposing. A variety of other state and federal privacy laws and regulations can present overlapping considerations.

§ 66:4.7 Location

Location data, particularly from mobile devices (mobile phones first among them, but also a plethora of “wearable” devices), are highly valued. Location data are collected from a variety of sources, including IP addresses, GPS, cellphone receivers, and other sensors. Many apps capture and sell location data generated by their users.

Location data is widely considered to be a sensitive data category under privacy laws at the state, federal, and international level. The collection and use of precise geolocation information, in particular, is subject to heightened scrutiny. For example, the U.S. Federal Trade Commission (FTC) has said that it considers precise personal location data to be “sensitive data” that warrant protection at a level akin to personal medical information, and states an expectation that companies provide transparency and choice about the collection and use of this information.¹⁹ State Attorneys General also have taken an interest in protection of location data, typically under state unfair and deceptive practice laws (known as UDAP statutes), and generally track the FTC guidance on expectations for companies in their enforcement actions.²⁰

Laws are evolving to provide explicit protection to location information. The CCPA specifically includes location data as a category of regulated “personal information,” and a number of jurisdictions are considering bans on the sale of location data.²¹ Generally, more

19. This view was established in a landmark 2012 FTC report on privacy. PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE, FTC REPORT (Mar. 2012), <https://www.ftc.gov/news-events/press-releases/2012/03/ftc-issues-final-commission-report-protecting-consumer-privacy>.

20. For a comprehensive overview of the law around location data, see Paige M. Boshell, *The Power of Place: Geolocation Tracking and Privacy*, BUS. L. TODAY (Mar. 25, 2019), <https://businesslawtoday.org/2019/03/power-place-geolocation-tracking-privacy/>.

21. For discussion of a particular local law initiative, see Jeffrey C. Bays, *New York City to Consider Banning Sale of Cellphone Location Data*, N.Y. TIMES (July 23, 2019), <https://www.nytimes.com/2019/07/23/nyregion/cellphone-tracking-location-data.html>.

restrictive norms for transparency and choice over the collection of location data continue to develop, and even when such datasets are offered in an “anonymized” form, increasing evidence casts doubt on the deidentified nature of large location datasets.²² The legal risk and rapid shifts in consumer expectations around location data thus make these datasets—while possibly some of the most useful information—at among the highest risk.

§ 66:4.8 Biometric

An emerging data source is biometric information captured by various technologies including AI facial recognition. Again, this can be highly sensitive.

In addition to the application of general privacy laws and similar regulatory scrutiny provided to other sensitive data like location information, several states have passed biometric privacy laws that can add a particular layer of risk.²³ Illinois's Biometric Information Privacy Act (BIPA) is considered the leading edge of the law in this

-
22. See, e.g., Jennifer Valentino-Devries, Natasha Singer, Michael H. Keller & Aaron Krolik, *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), <https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html>.
 23. Susan Crawford, *Facial Recognition Laws Are (Literally) All Over the Map*, WIRED (Dec. 16, 2019), <https://www.wired.com/story/facial-recognition-laws-are-literally-all-over-the-map/>. The opening lines of Ms. Crawford's article immediately illustrate the fractured landscape:

The current state of rules for use of facial recognition technology is literally all over the map. Next month, the city council in Portland, Oregon will hold a public meeting about blocking use of the technology by private companies, as well as by the government. San Francisco, Oakland, California, and Somerville, Massachusetts, already have banned the use of facial recognition technology by city agencies; Seattle's police stopped using it last year; and Detroit has said facial recognition can be used only in connection with investigation of violent crimes and home invasions (and not in real time). State governments have their own rules too. In October, California joined New Hampshire and Oregon in prohibiting law enforcement from using facial recognition and other biometric tracking technology in body cameras. Illinois passed a law that permits individuals to sue over the collection and use of a range of biometric data, including fingerprints and retinal scans as well as facial recognition technology. Washington and Texas have laws similar to the one in Illinois, but don't allow for private suits.

See also Tom Taulli, *Facial Recognition Bans: What Do They Mean for AI (Artificial Intelligence)*, FORBES (June 13, 2020), <https://www.forbes.com/sites/tomtaulli/2020/06/13/facial-recognition-bans-what-do-they-mean-for-ai-artificial-intelligence/#5ff99c5846ee>.

area and, with its private right of action and statutory damages, has inspired expensive litigation for companies that collect and maintain biometric identifiers. These laws increase the risk of holding information that might be subject to a biometric privacy requirement and also increase the notice and consent burdens on the process of collecting this information, which in turn adds to the diligence needs for any consumers of such datasets.

In addition to traditional personal privacy considerations, biometric data raise public policy questions, especially around surveillance, security, bias, and data integrity. Accordingly, debate about the ethical use of these technologies is raging. These debates may impact the future availability of biometric datasets and should be a consideration in a firm's risk-based data strategy.

§ 66:5 U.S. Legal Principles and Regulatory Views and Trends

We start with some basic table-setting. The primary regulator for the U.S. investment management sector is the Securities and Exchange Commission (SEC), and the SEC's core expectations are that:

- (1) An investment manager should fully disclose risks, fees, expenses, and conflicts of interest to which the manager's clients will be subject; and
- (2) An investment manager should maintain internal policies and procedures reasonably designed to identify and manage those risks and conflicts of interest, including the operational risks implicit in the investment manager's business.²⁴

The SEC, for the most part, does not regulate an investment manager's activities in advance. Instead, regulation is *ex ante*, with SEC staff inspecting an investment manager's business over time and posing questions about business practices and the sufficiency of risk disclosure and mitigation after the fact.

The SEC, also for the most part, does not directly regulate data practices for investment managers, with the most salient examples of direct regulation by the SEC being limited to its administration of Regulation S-P and attention to cybersecurity and data safeguarding practices. Most sources of regulation associated with new data and AI initiatives thus will be outside the traditional scope looked to by investment management lawyers and compliance professionals.

24. These are understood as extensions or manifestations of an investment adviser's fiduciary duty owed to its clients, which encompass duties of loyalty and care.

It is with these principles in mind that the “patchwork regulation” of data practices for investment managers should be considered.²⁵ Multiple sources of law must be identified and understood. Then, a firm’s legal and compliance practitioners should think broadly about each from the perspectives of “what risks and conflicts of interest does this present for the firm and its clients,” “how can the risks and conflicts be mitigated,” and “assuming they are material, what disclosures should be made.”

Building from there, the remainder of this chapter is divided—admittedly somewhat arbitrarily—into two sections: first, this section 66:5, which provides an overview of relevant legal principles, and then section 66:6, which outlines compliance considerations that follow from those principles. Data sourcing practices—a core data-related consideration for investment management legal and compliance professionals—are covered in section 66:6.

Likely sources of legal risk associated with an investment management firm’s data initiatives are discussed in the sections below.

§ 66:5.1 Data Protection and Cybersecurity

A threshold operational and legal risk presented by data—whether self-generated or acquired from someone else—is the need to protect the data. A data protection strategy will be driven by a variety of related goals, namely:

- (1) Avoiding literal loss, theft, or corruption;
- (2) Establishing protections against third-party infringement; and
- (3) Compliance with specific legal requirements attaching to the data.

[A] Avoiding Loss

Safeguarding sensitive digital data is central to every firm’s cybersecurity efforts. Reasonably extensive guidance is available discussing SEC expectations for investment manager cybersecurity policies. These include both survey findings from the SEC office that inspects investment managers (the Office of Compliance Examinations or Inspections, or OCIE) and guidance that can be intuited from the allegations in the multiple instances when the SEC has sued firms

25. Patchwork regulation is a widely used term. Other terms intended to capture much the same concept are “sectoral regulation” and “horizontal regulation”—the idea being in each case that the sources of regulation are diffuse and, in the case of data, often vary by type, use-case and jurisdiction.

for alleged cybersecurity deficiencies.²⁶ Information security is also on OCIE's announced list of 2020 "examination priorities" (as it was in prior years).²⁷

To date, SEC enforcement has focused on breaches involving exposure of client or customer personal information; there has been less interest in other, more general cybersecurity concerns for the investment management industry, such as protection of core operating systems or "crown jewels"-type information, ransomware deterrence and the like. That said, a worry for any organization that expands its data profile—even when the data involved may not be personal data of the type that has generated SEC enforcement—is the possibility that the organization may be a more attractive target for a hack, intentional sabotage of its data or data sources, corporate espionage, or other attack.

[B] Protection Against Infringement

Infringement here refers not to a specific term of art, but to a bundle of concepts akin to virtual trespassing or virtual appropriation. This builds on the reality that who generates, owns, and controls data is not always clear. Thus, a thoughtful data strategy will focus in part on better delineating ownership and control.

In the first instance, this typically will be a question of contract. For a data-aware firm, detailed and thoughtful data protection and ownership terms need to be considered in connection with potentially every vendor contract, every customer or client contract, a firm's terms of employment, and employee manuals, its website terms of use pages, and so on. And, of course, this is a two-way street. At the same time as the data-aware firm is ring-fencing its data contractually, its service providers and counterparties (likewise data-aware) are

26. OCIE has published multiple cybersecurity "risk alerts" after conducting a series of targeted industry examinations on the topic. *See* OCIE's 2015 Cybersecurity Examination Initiative (Sept. 15, 2015), <https://www.sec.gov/ocie/announcement/ocie-2015-cybersecurity-examination-initiative.pdf>; U.S. Sec. & Exch. Comm'n, Observations from Cybersecurity Examinations (Aug. 7, 2017), <https://www.sec.gov/files/observations-from-cybersecurity-examinations.pdf>; U.S. Sec. & Exch. Comm'n, Cybersecurity and Resilience Observations (Jan. 27, 2020), <https://www.sec.gov/files/OCIE%20Cybersecurity%20and%20Resiliency%20Observations.pdf>.

Other financial regulators also have published extensive guidance for cybersecurity standards, including the Financial Industry Regulatory Authority (FINRA), Consumer Financial Protection Bureau (CFPB), Federal Financial Institutions Examinations Council (FFIEC), and the New York State Department of Financial Services (DFS).

27. OCIE 2020 Examination Priorities (Jan. 2020), <https://www.sec.gov/about/offices/ocie/national-examination-program-priorities-2020.pdf>.

nibbling at the perimeter, laying claim to data generated as part of their relationships with the firm. Protecting a firm's data thus means being sure that any third-party claim is consistent with the firm's view of the relationship.

There also may be intellectual property bases to protect data, notably as a trade secret or other proprietary and confidential information protected by contract. While data generally cannot be patented or copyrighted, systems for analyzing data, especially if grounded in technology, might be patentable, and how a database is arranged, organized, and presented might be protected by copyright.

§ 66:5.2 *Legally Protected Data—Privacy*

The type of data most likely to carry a legal requirement to protect is personal data associated with individuals, especially names, addresses, government identification numbers, and the like. The definition of what data constitutes personal data under privacy laws has greatly expanded in recent year, with U.S. laws increasingly adopting a more European view of personal data as anything that reasonably is capable of being associated with a natural person—and sometimes even a device. Accordingly, lawyers and compliance professionals should be especially focused on the risks presented for their organizations in handling this type of data.

As described at sections 66:3 and 66:4 above, a variety of data gathered and used by investment managers presents privacy and data protection law considerations. That said, many investment managers also make the prudential decision, when practicable, to explicitly avoid collecting personal data. That is at least when speaking about data used for investment research; such a judgment is harder to live with in the context of data used in a firm's marketing or client service initiatives.

At the leading edge of comprehensive personal privacy laws are the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), each summarized below. Both are relatively new, with the GDPR taking effect in 2018 and the CCPA in 2020. Various U.S. state and federal laws, or heightened standards under international laws, also separately address special protections for populations deemed especially vulnerable, such as children, or sensitive personal information tied to health and financial records, gender orientation information, and political and religious affiliations, as well as privacy laws focused on certain technologies. Examples include the financial privacy laws and regulations cited earlier in the chapter (for example, the Bank Secrecy Act, Gramm-Leach-Bliley Act and its Regulation S-P, and the Fair Credit Reporting Act), health privacy laws such as the Health Insurance Portability and Accountability Act (HIPAA), or privacy

laws protecting certain content and other usage information such as the Electronic Communications Privacy Act (ECPA) and the Video Privacy Protection Act (VPPA). These more specialized laws and regulations will be critically important to review and understand when the nature of an investment manager's data practices is within their scope, but they generally are not treated in any detail in this chapter.

The risk of fines and penalties associated with a breach of the privacy and data protection requirements under these and other data protection laws is not insignificant and can include specified statutory damages. Reputational harm can be even more damaging, especially with the viral publicity that data-related cases can receive.

[A] GDPR

The GDPR, which became effective in 2018, generally is considered the most comprehensive data protection framework worldwide and is covered here at a high level notwithstanding that it is not a U.S. law or regulation. An investment manager can find itself subject to the GDPR if it is established or has an establishment in the European Union, endeavors to offer goods and services to individuals in the European Union, or monitors the behavior of individuals in the European Union. An investment manager also may be contractually subjected to requirements stemming from the GDPR as a result of global business relationships, even if it is not directly subject to the jurisdiction of the GDPR.²⁸

There are specialized requirements that apply depending on the type of processing an investment manager undertakes, but this level of detail is beyond the scope of this chapter. Broadly speaking, investment managers would be required, under the GDPR, to determine and define their status under the GDPR as a controller or processor for the data it processes, comply with a variety of administrative compliance requirements, and ultimately to process personal data for a fair, specified, and legitimate purpose, among a handful of legal bases under the regulation.²⁹ The information must be protected by appropriate technical and organizational security measures, and not stored for any longer than required for its stated purpose.³⁰ Stricter rules attach to certain categories of data, which are deemed by their nature sensitive. This includes information relating to an individual's criminal records, health, race, political opinions, religion, etc.³¹ Additionally, the GDPR requires transparency and sometimes very detailed disclosures about the processing of personal data, as well as

28. Arts. 3–4, Regulation (EU) 2016/679 (GDPR).

29. Art. 6, GDPR.

30. Art. 5, GDPR.

31. Arts. 9–10, GDPR.

the facilitation of a variety of data subject rights, including the right to request deletion of data or that personal information no longer be processed. Only data relating to “identifiable natural persons in the EU” are in scope to the GDPR, such that deidentified data can be obtained and processed without being subject to the GDPR. As noted below, pseudonymized data are still subject to the law.

Even non-EU investment managers with purely domestic operations and relationships also may find themselves indirectly affected by the operation of the GDPR, as the regulation prohibits the transfer of personal data outside of the European Economic Area (EEA) even when transferred to third parties within the same group of companies, unless there is a lawful basis for the transfer. One basis for transfer outside of the EEA is for a transfer to be to a jurisdiction that is deemed to provide an adequate level of data protection essentially equivalent to the European Union. A handful of jurisdictions have obtained such adequacy determinations, including Canada, Israel, and Argentina. The United States has not. Particularly in the wake of a July 2020 decision of the Court of Justice of the European Union that invalidated the Department of Commerce EU-U.S. Privacy Shield Framework, the most common basis for transfers to the United States is the execution of standard contractual clauses (SCCs), also sometimes called model contracts.³²

[B] CCPA

The U.S. state law drawing the most attention at present is the California Consumer Privacy Act (CCPA), which became effective January 1, 2020, and became enforceable by the California Attorney General on July 1, 2020. There are also a set of implementing regulations by the California Attorney General that have had and will continue to have evolving impacts on the data economy. Moreover, in addition to several potential amendments, the CCPA may be replaced by a California ballot initiative called the California Privacy Rights Act (CPRA), which has qualified for the November 2020 ballot. The CCPA—and potentially, the CPRA—force further transparency and consumer choice over the collection, use, and especially sale of personal information, and also add to the legal risk and compliance burden for firms that amass large volumes of data.

The CCPA requires new and heightened disclosures about the personal information that is collected, maintained, shared, and sold. The law applies to for profit entities that collect personal information of California residents, do business in California, and meet one of three thresholds: annual gross revenues in excess of \$25 million; annually

32. Arts. 44–45, GDPR.

buys, receives, sells or shares for commercial purposes the personal information of 50,000 or more consumers, households or devices; or derives 50% or more annual revenue from the sale of personal information.³³ Because nearly every U.S. business has California touch points, the law has the potential to set *de facto* new national standards.

The CCPA establishes specific rights of California consumers, which can be summarized as follows:

- (1) The right to know what consumer personal information is collected by businesses;
- (2) The right to know whether the personal information is sold or disclosed, and to whom such information is sold or disclosed;
- (3) The right to opt out of the sale of personal information;
- (4) The right to access the personal information (both categorically and the “specific pieces” of data);
- (5) The right to request that personal information be deleted; and
- (6) The right to non-discrimination, meaning equal service and price, even if privacy rights are invoked.

A wide range of consumer personal data is protected. The law defines personal data in part as “information that identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked with a particular consumer or household.” This includes, but is not limited to:

- (1) Identifiers (for example, name, contact information, ID numbers, IP address);
- (2) Personal information under the California disposal law (for example, name and SSN);
- (3) Characteristics of protected classifications under California and federal law (for example, race and gender);
- (4) Commercial information (for example, purchasing histories or tendencies);
- (5) Biometric data;
- (6) Internet or other electronic network activity information (for example, browsing and search history and interactions with websites, apps, and ads);
- (7) Geolocation data;

33. CAL. CIV. CODE § 1798.140(c).

- (8) Audio, electronic, visual, thermal, olfactory, or similar information;
- (9) Professional or employment-related information;
- (10) Education information (as defined in the Family Educational Rights and Privacy Act (FERPA)); and
- (11) Inferences drawn from any of the above to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitude.

Personal information subject to the CCPA does not include information lawfully made available from federal, state, or local government records, as well as deidentified or aggregated and anonymized information.

[C] A Note on Deidentified or Aggregated Data

Deidentified and aggregated (anonymized) data are commonly acquired and are understood to be outside many privacy laws. However, whether anonymization has been effective depends on the facts and can be questioned.³⁴ Deidentification standards also vary depending on the regulatory regime that applies to the personal information in question.

For example, the CCPA includes specific requirements for deidentification. First, the data must no longer be “information that identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked with a particular consumer or household” (a circular statement because this simply says the information is no longer personal information within the meaning of the law). Second, the business claiming deidentification protections must:

- (1) Implement technical safeguards that prohibit re-identification of the consumer to whom the information may pertain;
- (2) Implement business processes that specifically prohibit re-identification of the information;

34. The effectiveness of anonymization under the GDPR was subject to almost immediate debate, as described in a 2018 report by the advocacy group Privacy International, see PRIVACY INT'L, WHY WE'VE FILED COMPLAINTS AGAINST COMPANIES THAT MOST PEOPLE HAVE NEVER HEARD OF—AND WHAT NEEDS TO HAPPEN NEXT (Nov. 8, 2018), <https://privacyinternational.org/advocacy/2434/why-weve-filed-complaints-against-companies-most-people-have-never-heard-and-what>. See also Aliya Ram & Madhumita Murgia, *Data Brokers: Regulators Try to Rein in the 'Privacy Deathstars,'* FIN. TIMES (Jan. 10, 2019), <https://www.ft.com/content/f1590694-fe68-11e8-aebf-99e208d3e521>.

- (3) Implement business processes to prevent inadvertent release of deidentified information; and
- (4) Make no attempt to re-identify the information.

Another key standard for deidentification stems from HIPAA. Protected Health Information (PHI) subject to HIPAA may be considered deidentified only if either:

- (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information and documents the methods and results of the analysis that justify such determination; or
- (2) Specified identifiers of the individual or of relatives, employers, or household members of the individual, are removed.³⁵

Once PHI is deidentified, the requirements of the HIPAA Privacy Rule no longer apply, except that disclosure of a re-identification code or methodology would then itself constitute potentially unlawful disclosure of the PHI and, if the information is at any time re-identified, it becomes subject again to PHI protections and restrictions.³⁶

Pseudonymized data, in which the data are still presented as relating to a specific individual or household, but a pseudonym or similar mask has been applied to conceal identity, is a variation on anonymization processes. It is favored as a privacy protective processing safeguard, but because pseudonymized data still are tied to an individual identifier and allow for granular data analysis, it does not necessarily take the data outside of the scope of the privacy laws. Indeed, pseudonymized data are a separate category of data subject to the GDPR, and attempts to amend the CCPA to clarify that pseudonymized data are not personal information under the law were unsuccessful.

Investment management firms that accept deidentified, and/or pseudonymized data, generally have internal requirements instructing staff not to attempt to unmask or re-identify the data, which requires care about data appending, or combinations of datasets that may result in inadvertent unmasking or re-identification. These processes to preserve the deidentification or pseudonymization of the

35. 45 C.F.R. § 164.514(b).

36. 45 C.F.R. § 164.502(d)(2).

data support risk mitigation from privacy laws, and are often a contractual requirement from the provider of the relevant datasets. These administrative and procedural safeguards, along with other data governance practices, are further explored in 66:5:3 below.

§ 66:5.3 *Select Additional Examples of Protected Data—Government Data, National Security Data, and the CFAA*

[A] *Government Data*

There is a natural presumption that governmental data, especially in democratic societies, is intended to be “open” and freely accessible to the public. In fact, despite the U.S. commitment to data accessibility, including as mandated by President Trump’s recent AI Executive Order, this is not guaranteed, and certain kinds of data remain protected and withheld from the public. Moreover, permitted uses of governmental data, even when the data can be readily accessed, can be context specific. For example, some public data sources may be presented with the disclaimer that they are intended for research and other non-commercial purposes. There are also a variety of instances when governmental data might be explicitly non-public and restricted, for example, in connection with governmental contracts, studies, and approvals that have not yet been announced.

The tension between those principles—“open government” versus access and use restrictions—is illustrated in various ways. As a modest example, there have been claims brought against the U.S. federal courts for charging fees for court records; claimants argue the fees infringe on their right to access public information.³⁷ The federal Freedom of Information Act (FOIA) and its various state “sunshine law” analogs also illustrate the tension between open government principles and the reality that many categories of information collected by the government must be kept confidential. FOIA starts from the presumption that governmental data should be made available on request, but then establishes categories of information that are exempt from disclosure.³⁸

37. For a description of the cases and the legal and policy arguments, see The Editorial Bd., *Public Records Belong to the Public*, N.Y. TIMES (Feb. 7, 2019), <https://www.nytimes.com/2019/02/07/opinion/pacer-court-records.html>.

38. The nine categories of FOIA-exempt information are: national security information; information relating solely to the internal personnel rules and practices of an agency; information prohibited from disclosure by another, superseding federal law; trade secrets or confidential commercial information; privileged communications within or between agencies;

The SEC and U.S. Department of Justice have brought cases involving so-called “political intelligence” operations, generally referring to the collection of government information before it is widely disseminated. The first high-profile case resulted in a settlement with the SEC in which a political intelligence firm agreed to enhance its policies and procedures for handling sensitive government information.³⁹ In a later case, the Department of Justice prosecuted and convicted four individuals—a government insider, a political intelligence consultant, and two portfolio managers at an investment firm—for alleged insider trading involving information from a government health insurance rate-setting office regarding upcoming reimbursement rule changes.⁴⁰

[B] National Security

Data that may qualify as protected under classifications for national security purposes are also higher risk, and the dividing line between national security and commercial considerations is increasingly blurred. Outside the United States, national security data can be especially difficult to divide from commercial data. As an example that could give rise to concern in any country, imagine a data collection program gathering public information on critical infrastructure such as dams, power plants, and the like; imagine further that the data are then transferred outside the host country. Such a program may be entirely innocent but still could be misconstrued and generate national security concerns and governmental investigation. (In the United States, information related to critical infrastructure is protected as confidential by a variety of governmental programs; private entities in those sectors must be cognizant of potential restrictions in the dissemination of operational data.) National security issues are magnified in jurisdictions with significant state ownership of what otherwise would appear to be traditional commercial enterprises, and certain international cybersecurity laws have expansive requirements that bring large sectors of critical infrastructure industries within the scope of government oversight and even pre-approval programs.

personal privacy information; law enforcement information; information concerning the supervision of financial information; and geological information on wells. Further details on FOIA, see FOIA, <https://www.foia.gov/faq.html>.

39. *In re Marwood Group Research, LLC*, SEC Release No. IA-4279 (Nov. 24, 2015), <https://www.sec.gov/litigation/admin/2015/34-76512.pdf>.

40. Christian Berthelsen & Bob Van Voris, ‘*King of Political Intelligence*’ Sentenced to Prison for Insider Trading, BLOOMBERG (Sept. 13, 2018), <https://www.bloomberg.com/news/articles/2018-09-13-king-of-political-intelligence-gets-one-year-in-insider-case>. (The convictions are on appeal.)

[C] Web Scraping and the Computer Fraud and Abuse Act (CFAA)

The most common form of alternative data developed or purchased by investment managers today is web-scraped data. As described above, web scraping, also referred to as crawling or spidering depending on the technologies used, is the automated gathering of data from a third-party website. Scraping is critical to business processes across many industries and is therefore in wide use. But the permissibility of the practice—and associated legal risks—remain unclear. A variety of legal claims may apply under U.S. law to unauthorized scraping, including breach of contract, copyright infringement, trespass and other torts, and state and federal laws specific to website access.

Federal law (the Computer Fraud and Abuse Act or CFAA), enforceable both criminally and civilly, specifically protects websites from unauthorized access, with that phrase potentially extending the law's protections to any website whose terms of use forbid or limit automated scraping of data from the site.⁴¹ Applicability of the CFAA is perhaps the most debated legal topic associated with scraping and a matter of keen interest to investment managers consuming scraped data. A key question in the circuit splits and debate is whether the law reaches access to and manipulation of data within a protected computer for a *purpose* that was unauthorized, or simply prohibits access in the first instance that was unauthorized. In addition to the impact on web-scraping activities, the resolution of the CFAA interpretation also has significant consequences for theft and misuse of trade secrets and other insider risk management for employees who violate company policy in their use of IT resources.

A widely cited case on point, frequently considered in the web-scraping context, is that of *HiQ Labs Inc. v. LinkedIn Corp.*, which arises from claims by LinkedIn that HiQ violated the CFAA by scraping LinkedIn user pages in contravention of a formal cease-and-desist notice. While the case is ongoing and subject to appeal to the U.S. Supreme Court, the decisions to date, including a decision by the 9th Circuit Court of Appeals,⁴² generally favor HiQ. The decisions generally are being read as supportive of the view that commercial scraping of publicly available data, when the scraping does not interfere with a website's operations, is not subject to the CFAA. However, similar cases that could be read to mitigate risk from web scraping have not been decided in every jurisdiction, so risk remains, at least until the law is clarified by the Supreme Court or legislative process.

41. The CFAA makes it unlawful to “intentionally access” a computer or website without authorization or in a manner that “exceeds authorized access.” There also are numerous U.S. state law analogs to the CFAA.

42. *HiQ Labs Inc. v. LinkedIn Corp.*, 938 F.3d 985 (9th Cir. 2019).

§ 66:5.4 Insider Trading

Another top-level risk to be considered is that associated with insider trading laws, and in particular, the possibility that data obtained by an investment manager might incorporate material non-public information (MNPI). A ready example of inside information is a prospective large-scale transaction of XYZ Corp. kept under wraps: the information relates to XYZ Corp., it is specific, it has not been made public and, upon becoming public, it likely will affect the price of XYZ Corp. shares.

Big data do not slot as neatly within the definition of inside information or MNPI. Consider, for example, a bank's credit card transaction data. Might the data represent inside information or MNPI *as to the bank* by revealing the volume of credit card transactions the bank is handling? Or might the data represent inside information or MNPI *as to a particular retailer* by revealing sales information before it can be aggregated and publicly disseminated by the retailer? This uncertainty, while an interesting intellectual exercise, does in fact entail significant risk for investment managers, who may face both civil and criminal consequences for being in breach of insider trading laws.

The latter risk—that consumer transaction data might be MNPI vis-à-vis a retailer—is highlighted by an insider trading case brought by the SEC against a bank employee (a fraud detection analyst) who, in the ordinary course of business, had access to real-time information on credit card transactions processed by the bank. The employee allegedly developed a software program based on the data that permitted him to extrapolate a retailer's overall sales figures and then trade in the securities of that retailer when his program predicted the retailer's sales would vary from its publicly reported forecasts (for example, disappoint or positively surprise the market). Among other things, the employee argued that the bank saw only a very small percentage of a given retailer's credit card transactions, a basis to claim that the data he had was non-material. The court rejected that defense and accepted the premise that credit card transaction data, on those facts and when used in that manner, constituted MNPI.⁴³

The typical issue faced by an investment manager is, of course, much more nuanced. For example, assume that fluctuations in a company's hiring activity might influence a trader's decision whether to buy or sell the company's securities. On that basis, a traditional word of mouth insider "tip" ("just heard that XYZ Corp. pulled its recruiting searches, could mean they're not growing anymore") might be readily understood as carrying potential risk. But what about when

43. SEC v. Huang, No. 16-2390 (3d Cir. 2017), <https://law.justia.com/cases/federal/appellate-courts/ca3/16-2390/16-2390-2017-04-10.html>.

that same fact—that XYZ Corp. pulled recruiting searches—instead can be divined from a mountain of job and recruiting search data housed by or visible on online jobs websites? Because there is no “tip,” and the corresponding information might be obscured or aggregated within a larger dataset covering many companies, it is understandably less likely to set the same alarm bells ringing.

To help guard against this MNPI risk, sophisticated investment management consumers of data will ask questions intended to confirm the data were legitimately obtained without any violation of a duty of confidentiality or loyalty along the way. MNPI risk associated with potentially “misappropriated” data is an important driver of an investment manager’s controls associated with data sourcing, including web-scraping controls, as discussed at section 66:6 below.⁴⁴

§ 66:5.5 Artificial Intelligence Techniques

Regulatory views on AI are for the most part early stage and evolving, but recent years have seen an upswing in pronouncements.

[A] U.S. Treasury Report

One of the broadest and most comprehensive discussions specific to AI in financial services is a 2018 report prepared by the U.S. Treasury Department.⁴⁵ The report opens by observing that AI investment by financial services firms is accelerating and that AI innovations are driving efficiencies for firms and improved outcomes and choices for their customers. The report expresses concern, however, that “black box” systems are inconsistent with traditional regulatory expectations of transparency and auditability for industry activities. Indeed, the concern that AI tools will lead to determinations that humans find difficult to audit or track the logic of, particularly when the determinations have a significant impact on individual rights, liberties, or even economics, are at the heart of many of AI critiques.

44. The possibility that web-scraped data might be viewed as “misappropriated” for this purpose has been significantly mitigated by the favorable *HiQ* case law to date (see discussion at text accompanying *supra* note 42), but still should be considered in the context of actual facts.

45. U.S. Dep’t of the Treas., Report to Pres. Donald J. Trump, A Financial System that Creates Opportunities: Nonbank Financials, Fintech and Innovation (July 2018), https://home.treasury.gov/sites/default/files/2018-08/A-Financial-System-that-Creates-Economic-Opportunities---Nonbank-Financials-Fintech-and-Innovation_0.pdf. An earlier and also very detailed report is by the Financial Stability Board, a global regulatory group. See Press Release, Fin. Stability Bd., FSB Considers Financial Stability Implications of Artificial Intelligence and Machine Learning (Nov. 2017), <http://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/>.

Treasury also suggests that AI presents a variety of two-edged sword risks: for example, trading will become even faster and more efficient, but potentially at risk of new bouts of volatility; or new tools might help root out rogue traders, money launderers, cyber criminals, and other bad actors, but bad actors surely will challenge systems with their own sophisticated applications as well.

[B] FINRA Guidance

The U.S. Financial Industry Regulatory Authority (FINRA) has published detailed regulatory guidance for broker-dealers (an industry closely adjacent to investment management) in what is being called the FINRA AI 2020 Report.⁴⁶ The report is the culmination of a two-year review by FINRA's Office of Financial Innovation to learn about the emerging challenges confronted by broker-dealers and other market participants as they introduce AI-based applications into their businesses. The report covers similar ground to this chapter and provides an overview of AI technology, explores its diverse applications in the securities industry, and identifies challenges and legal considerations.

Among the challenges and legal considerations highlighted by FINRA are model risk management, data governance, customer privacy, and supervisory controls systems, as well as cybersecurity, outsourcing and vendor management, books and records requirements, and workforce structure. FINRA also emphasizes that this list is not exhaustive. Every firm should conduct its own due diligence prior to implementation to determine the utility, legal impact, and other potential risks of an AI-based application.

FINRA advises firms to incorporate in their policies and procedures a model validation process sufficient to respond to the intricacies of an AI-based application. Such a review would involve searches for potential biases in the input data and errors in the algorithm, verification of the risk thresholds, and an articulation of the explainability of the output. Assessments also might include how outputs are derived, whether those outputs align with the firm's business goals, risk appetites, internal policies and procedures, and whether the actions taken pursuant to those outputs comport with the firm's legal and compliance requirements.

FINRA notes the importance of robust data flows for AI applications and encourages firms to emphasize data source verification, data integration, data security, and data quality benchmarks and metrics. Firms also should enhance their policies and procedures to

46. FINRA AI 2020 Report, *supra* note 4.

include a review of the underlying datasets for any potential built-in biases. FINRA further notes that firms likely will collect (and therefore must appropriately safeguard) sensitive customer data such as personal identifiable information (PII) when employing AI-based applications. Policies and procedures should take into account issues such as customer consent, user entitlements and access to data, and the redaction of sensitive information.

[C] SEC Guidance

The SEC has not directly spoken to how investment managers and other SEC-regulated firms should consider their use of AI.⁴⁷ The agency has, however, brought a number of enforcement actions involving alleged failures by firms to properly vet and implement complex investment models, generally also alleging related failures to disclose weaknesses or limitations in the models.⁴⁸ It is predictable that this enforcement history will inform SEC thinking on use of AI by the agency's regulated firms. (SEC expectations are further reviewed at section 66:6 below in connection with recommended quantitative model governance practices.)⁴⁹

-
47. Various SEC personnel have spoken regularly about the SEC's own use of data and AI. *See, e.g.*, Kara M. Stein, SEC Comm'r, Speech at the Georgia State University College of Law—Henry J. Miller Distinguished Lecture Series: From the Data Rush to the Data Wars: A Data Revolution in Financial Markets (Sept. 27, 2018), <https://www.sec.gov/news/speech/speech-stein-092718>; Scott W. Bauguess, Acting Dir. & Acting Chief Econ., SEC Div. of Econ. & Risk Analysis, Keynote Address at OpRisk North America 2017: The Role of Big Data, Machine Learning, and AI in Assessing Risks: A Regulatory Perspective (June 21, 2017) [hereinafter Bauguess Speech], <https://www.sec.gov/news/speech/bauguess-big-data-ai>.
 48. See various SEC press releases (and the associated orders) describing these actions: SEC Press Release 2011-37, SEC Charges AXA Rosenberg Entities for Concealing Error in Quantitative Investment Model (Feb. 3, 2011), <https://www.sec.gov/news/press/2011/2011-37.htm>; SEC Press Release 2014-289, SEC Charges Investment Manager F-Squared and Former CEO with Making False Performance Claims (Dec. 22, 2014), <https://www.sec.gov/news/pressrelease/2014-289.html>; and SEC Press Release 2018-167, Transamerica Entities to Pay \$97 Million to Investors Relating to Errors in Quantitative Investment Models (Aug. 27, 2018), <https://www.sec.gov/news/press-release/2018-167>.
 49. While little has been said publicly in this direction, it also is reasonable to expect the SEC and SEC staff will require—for example, in the course of SEC inspection activity—that investment management firms be able to demonstrate and explain the basis of their AI-based or other quantitative investment decisions. This accords both with traditional books and records requirements applicable to investment decisions and, more generally, with an investment manager's fiduciary duty of care.

[D] U.S. Federal Reserve Guidance

Another widely cited source of regulatory guidance on AI in financial services came in a speech by a member of the board of governors of the U.S. Federal Reserve, who suggested “existing regulatory and supervisory guardrails”—and especially existing guidance on vendor oversight and diligence and risk management when using complex models—provide a sufficient starting point. In other words, new U.S. banking regulation specific to AI might be required in the future, but not yet.⁵⁰

[E] Source Code Access

U.S. investment management regulators have taken different tacks over time with respect to demanding access to sensitive source code when supervising businesses deploying AI or other sophisticated software applications. As an indication of how concerned some parties are that source code will be mishandled by the government (the highest order concern being that a company’s intellectual property might be stolen by hackers or even bad actors inside the government), Congress debated the Protection of Source Code Act, which would have prohibited the SEC from accessing source code at SEC regulated businesses without obtaining a subpoena. The House passed the bill, but it appears to have died in the Senate at the end of 2018.

The same issues animated the U.S. Commodity Futures Trading Commission (CFTC), which grappled with the question in the course of developing its Regulation AT (referring to automated trading). The regulation would have given the CFTC access to quantitative trading software source code at CFTC regulated firms, but opponents argued, first, that due process protections require a subpoena before access and, in any event, that the CFTC is ill equipped to protect sensitive intellectual property from loss or theft.⁵¹ The CFTC abandoned the initiative when Republican appointees became the majority on the CFTC following the election of President Trump.⁵²

50. Lael Brainard, U.S. Fed. Reserve Bd. of Governors, Speech at Fintech and the New Financial Landscape: What Are We Learning about Artificial Intelligence in Financial Services? (Nov. 13, 2018) [hereinafter Brainard Speech] [citing SR Letter 11-7 and SR 13-19/CA 13-21], <https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm>.

51. J. Christopher Giancarlo, Comm’r, Commodity Futures Trading Comm’n, Statement of Dissent Regarding Supplemental Notice of Proposed Rulemaking on Regulation Automated Trading (Nov. 4, 2016), <https://www.cftc.gov/PressRoom/SpeechesTestimony/giancarlostatement110416>.

52. U.S. Regulator Declares ‘Dead’ Move to Seize HFT Code, FIN. TIMES (Oct. 4, 2017), <https://www.ft.com/content/068ce050-a922-11e7-93c5-648314d2c72c>.

§ 66:5.6 Antitrust

There has been debate in recent years whether big data should have a special place in antitrust analysis. Some contend control of valuable datasets creates potentially new risk to market competition.⁵³ Others maintain data are like any other assets and should be analyzed under traditional antitrust analysis and as one piece of an overall understanding of a company's position in the competitive landscape.⁵⁴ In any event, given generally intense competition, the prospect that any one investment manager's access to data could represent an impermissible barrier to entry appears to be a purely hypothetical concern.

But there still can be antitrust risks associated with data, notably for private equity fund managers. Consider a private equity firm that owns several hotel chains. Occupancy statistics, room rate data, discounting strategies, customer lists, and the like associated with each hotel company generally are understood as sensitive information from an antitrust perspective. Such information is risky to share between competitors, given the possibility it might facilitate unfair collusion.⁵⁵ It follows that when such information is shared with a hotel company's upstream private equity owner, it is then incumbent on the private equity firm not to serve as a conduit through which the information is transmitted out to the company's competitors (including another hotel company in which the firm has an interest) without careful analysis of the associated antitrust risk.

-
53. TERRELL MCSWEENEY & BRIAN O'DEA DATA, U.S. FED. TRADE COMM'N, INNOVATION AND POTENTIAL COMPETITION IN DIGITAL MARKETS—LOOKING BEYOND SHORT-TERM PRICE EFFECTS IN MERGER ANALYSIS (Feb. 2018), https://www.ftc.gov/system/files/documents/public_statements/1321373/cpi-mcsweeney-odea.pdf (Ms. McSweeney is a former FTC Commissioner).

See also views expressed by European antitrust officials. For example, Margrethe Vestager, European Union Commissioner for Competition, has said data is becoming a “vital resource” and “competition can't work if just a few companies control a vital resource that you need to be able to compete—and if they refuse to share it with others.” Margrethe Vestager, EU Comm'r for Competition Speech at the Mackenzie Stuart Lecture, Cambridge: Making the Data Revolution Work for Us (Feb. 4, 2019), https://wayback.archive-it.org/12090/20191129203859/https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/making-data-revolution-work-us_en.

54. John Yun, *Antitrust After Big Data*, CRITERION J. INNOVATION (2019), <https://www.criterioninnovation.com/articles/antitrust-after-big-data/>; Timothy Muris & Jonathan Nuechterlein, *Antitrust in the Internet Era: The Legacy of United States v. A&P*, George Mason L. & Econ. Research Paper (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3186569.
55. Michael Bloom, *Information Exchange: Be Reasonable*, FTC BLOG (Dec. 11, 2014), <https://www.ftc.gov/news-events/blogs/competition-matters/2014/12/information-exchange-be-reasonable>.

This principle may be understood generally. However, as with the discussion of MNPI above, the prospect that competitive information might be embedded in or capable of being derived from big data—rather than in files plainly labeled as “occupancy statistics” or “discounting strategies”—creates a heightened risk of missteps.

§ 66:5.7 **Bias**

It is well documented that limitations in data—and notably using data collected mainly from majority or privileged populations—can embed or reinforce social and other forms of bias. There is clear policymaker and regulatory interest in the topic, as evidenced by statements requiring consideration of issues of data and bias in financial products. This notice from New York’s Department of Financial Services, which addresses life insurance, is an example:

An insurer should not use an external data source, algorithm, or predictive model for underwriting or rating purposes unless the insurer can establish that the data source does not use and is not based in any way on race, color, creed, national origin, status as a victim of domestic violence, past lawful travel, or sexual orientation in any manner, or any other protected class.⁵⁶

FINRA’s AI 2020 report also speaks to bias concerns. FINRA suggests a regulated broker-dealer risks not meeting its ethical obligations under FINRA Rule 2010 (a bedrock FINRA rule requiring member firms to act with high standards of commercial honor) if it disregards the potential for inherent biases in underlying datasets used in the course of the firm’s business.⁵⁷

Lawmakers on Capitol Hill are also increasingly looking at the impact of bias in financial services, including holding several hearings

56. N.Y. Dep’t Fin’l Servs., 2019 Circular Letter No. 1.

57. Other examples of federal financial regulators discussing the prospect of embedded bias in data include Lael Brainard’s speech, *see* Brainard Speech, *supra* note 50; a U.S. Treasury report, *see* U.S. Dep’t of Treas., Opportunities and Challenges in Online Marketplace Lending (May 10, 2016), https://www.treasury.gov/connect/blog/Documents/Opportunities_and_Challenges_in_Online_Marketplace_Lending_white_paper.pdf; and a no-action letter issued by the Consumer Financial Protection Bureau (CFPB) to a company that proposed to incorporate educational and other alternative data into its lending algorithm, with the letter conditioned on confirmation that the algorithm would be monitored to ensure it was not restricting access for underprivileged borrowers, *see* Press Release, Consumer Fin. Protection Bureau, CFPB Announces First No Action Letter to Upstart Networks (Sept. 24, 2017), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-announces-first-no-action-letter-upstart-network/>.

in the spring of 2020 on data bias, such as a hearing in the U.S. House Committee on Financial Services in February 2020 titled *Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services*.⁵⁸

In most data use cases for investment management, and certainly in the case of data used to inform investment decisions, the potential for discrimination and bias would appear to be a limited risk. An exception is the delivery of retail investment management advice, and especially robo-advice, where it is possible, for example, that the nature of the questions asked of customers during the on-boarding process could embed or reinforce bias.

§ 66:5.8 ***New York's Martin Act and General "Fairness" Principles***

The Martin Act, which a succession of New York Attorneys General have used to great effect in bringing securities fraud actions, generally prohibits "fraud" and "fraudulent practices" in connection with the offer, sale, or purchase of securities, but it differs from common law fraud. Common law fraud typically is understood as involving misrepresentations or omissions and to require scienter (or intent) to defraud. By contrast, the Martin Act does not require scienter. Courts instead have referred to its scope as prohibiting "deceitful practices contrary to the plain rules of common honesty."⁵⁹ Whether principles of "deceit" and "common honesty" should be stretched to address simple unfairness in the markets is, of course, doubtful. But the broad language of the Martin Act and a history of activist Attorneys General give pause. To date, there is no clarity on whether or how the Martin Act might be applied to big data- or AI-based investment strategies.

However, an example of the Martin Act being wielded in regard to data some years ago may warrant revisiting. In the case, the New York Attorney General opened an investigation into a news organization selling advanced access to economic survey data. Under pressure, the organization changed tack and set guidelines establishing more uniform access rules. While the New York Attorney General won the day, there were important voices on both sides of the debate, as evidenced by this commentary from a contemporaneous *New York Times* article:

58. See *Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services: Hearing Before the H. Comm. on Fin. Servs.*, 119th Cong. (Feb. 12, 2020), <https://financialservices.house.gov/calendar/eventsingle.aspx?EventID=406120>.

59. *People v. Fed. Radio Corp.*, 244 N.Y. 33, 38–39 (1926).

Legal standards aside, many market experts worry that the attorney general's crackdown may go too far and discourage the kind of enterprising research that has long made markets more efficient, whether undertaken by short-sellers hoping to profit directly on it, Wall Street investment banks that sell research to their clients, or investigative journalists. "The notion of a level playing field is important, and it's important to aim for it," said Harvey Pitt, a former SEC chairman and now the chief executive of the consulting firm Kalorama Partners. "But a level playing field can't mean everyone has the same information. People need financial incentives to dig up information, and the marketplace benefits."⁶⁰

§ 66:6 Data-Related Compliance and Controls Considerations

As with other aspects of an investment manager's compliance program and other internal controls, the first analytical step is often a risk inventory. The discussions above in this chapter are intended to inform that initial step. The next step is the development of thoughtful practices and controls that mitigate risk. Select areas for consideration are set out below.

§ 66:6.1 Data Sourcing

While some large datasets are sourced by investment managers directly (note especially the discussion of web-scraping compliance considerations above), the more common practice is to buy or license data from third parties. However data are sourced, the potential for legal liability is sufficiently high that robust controls should be established and maintained in both the selection of the data vendor, the contractual provisions respecting rights and responsibilities for the data, as well as the transfer and use of the data. Together with cybersecurity and other data protection protocols (noted in section 66:4 above), data sourcing controls are at the heart of an organization's efforts to mitigate its data-related risks. Disclosures of risks associated with alternative data are also beginning to appear.⁶¹

60. Nathaniel Popper, *Regulators Examining Sales of Early Financial Data*, N.Y. TIMES (July 8, 2013), <https://dealbook.nytimes.com/2013/07/08/regulators-examining-sales-of-early-financial-data/>; James B. Stewart, *Fair Play Measured in Slivers of Seconds*, N.Y. TIMES (July 12, 2013), <https://www.nytimes.com/2013/07/13/business/the-ethics-of-a-split-second-advantage-for-traders.html>.

61. While any disclosure should be thoughtfully tailored to the particular firm's circumstances, an example of risk disclosure relating to alternative data would be the following:

[A] Vendor Diligence

Investment managers routinely assess their data sellers from a compliance and risk management perspective. An investment manager purchasing data from a vendor wants to be sure the vendor is attuned to the same types of concerns that the manager has, that data lineage (discussed below) can be properly confirmed, that the vendor has obtained all rights and authorizations for the use and disclosure of the data in the proposed format (including authorizations from data subjects under applicable privacy laws), and that the vendor has some level of compliance infrastructure. Taken as a whole, the investment manager would like to be sure the vendor has an appropriate understanding and respect for both the various rules and contracts that might govern the vendor's rights in the data and the regulated context in which the investment manager is operating.⁶²

Investment and market analysis increasingly relies on inputs from so called "alternative data"—a phrase generally referring to information previously not available to professional investors because, for example, the implementing technology is relatively new (e.g., geolocation, equipment sensor, and "scraped" or "aggregated" website data), the information previously was held as proprietary (e.g., credit card processor, payroll, and shipping data), or the information was simply too unwieldy or expensive to manage. Risks associated with alternative data include the possibility of new legal and regulatory frameworks targeting collection, marketing and use of data (particularly relating to cybersecurity and privacy), technological changes that may make data more or less useful, changes in the marketplace for data (which is relatively new and fragmented), etc. For example, nearly all professional investors making regular use of alternative data rely on third-party data sellers or brokers and therefore are exposed to the possibility both that data may become unavailable or that it was improperly or illegally collected or handled, with derivative liability potentially extending to the investor purchasing the data. Insider trading and "fair practice" laws generally are untested in this sector and therefore present risks for alternative data-based investing. Holding sensitive data of any type also increases the risk of being targeted in various ways, e.g., by cyber-attackers, insider theft, government investigation or civil plaintiffs. Additional commercial risks relate to the prospect that investment decisions based on alternative data may be flawed for various reasons, including incomplete, "dirty" or misunderstood data and gaps or flaws in technology used to collect and analyze data. Substantial legal and technical costs may be incurred in responding to any of these risks and developments.

62. A working group at the Alternative Investment Management Association (AIMA) prepared a due diligence questionnaire (DDQ) tailored for use by investment managers with their data vendors. The DDQ is not publicly available, but bears some resemblance to DDQs in wide use with so-called expert network providers.

An investment manager also may wish to understand how the vendor's business practices more generally might present business and reputational risks. For example, some (but not all) firms prefer to interact only with vendors who offer data on a non-exclusive basis—meaning the same data purchased by one firm can be purchased by others on more or less the same terms. This interest in non-exclusivity tends to be driven by both fairness considerations like those outlined above in section 66:5's discussion of New York's Martin Act and the risk of receiving MNPI in an exclusive relationship.

[B] Data Lineage

Understanding “data lineage” (or “data provenance”) is critical to diligencing a dataset. The term refers to the concept that the purchaser or user of data should know enough about the chain of ownership to confirm the data was legitimately collected and appropriately managed and protected through the course of its existence. Understanding data lineage can be an important protection in mitigating insider trading risks (because data that is properly obtained and transferred over its lifecycle generally cannot be said to have been, using the rubric of U.S. insider trading law, “misappropriated”). In the ideal case, the investment manager trading on information derived from data will be able to confirm that the data were obtained legally and with third-party consents where applicable, that the further transfer of the data was likewise legal and consented to if required, and that disclosures associated with these permissions were appropriate and at least contemplate use of the data for commercial or business purposes, including sale.

[C] Personal Information

Many organizations take particular care in seeking to avoid the acquisition of personal data implicating privacy laws and regulations. Associated controls can include:

- (1) Specifications in contracts that the acquired data should not include personal data;
- (2) Diligence on the technical standards and processes used to create deidentified data in accordance with a generally acceptable standard for the context in which the data was derived;
- (3) Testing or auditing protocols designed to spot unintended inclusion of personal data;
- (4) Rules against re-identification efforts when handling anonymized or pseudonymized (which can include rules intended to prevent inadvertent re-identification, as can occur when multiple sources of information are cross-analyzed);

- (5) Rules against further disclosure or sale of the data to third parties or other requirements for contractual limitations on downstream recipient use or re-disclosure; and
- (6) Requirements that when personal data are nonetheless identified they will be embargoed and appropriately reported internally.

Where personal information is collected and maintained by the firm, these data flows must be considered and incorporated into the broader privacy and data protection compliance program, potentially impacting privacy policy disclosures, information security, the potential requirement to facilitate data subject rights, and employee training.

[D] Web Scraping/CFAA

Given the legal overlay reviewed above, investment managers that use scraped data often have compliance policies and procedures associated with web scraping. Elements of these policies vary, but can include:

- (1) Rules specific to scraping sites that include information on individuals (relevant to the determination that many investment managers make that they explicitly do not want to obtain personal data);
- (2) Requirements to honor cease-and-desist notices from website operators;
- (3) Prohibitions on efforts to circumvent password protection or technical barriers established by the website, including do-not-crawl signals;
- (4) Requirements to preclear or report password creation; and/or
- (5) Internal education and authorization requirements before staff can participate in scraping projects.

§ 66:6.2 Selling Data

Inevitably, some investment management firms will decide it is advantageous to sell some of the data that they assemble, or perhaps they will develop and market products that incorporate the data. This will require consideration of whether the original data acquisition activity and contracts contemplated later resale or incorporation of the data (resale may well be prohibited contractually or be outside the terms of the original consent). It also will require consideration of how such data sales activity is regulated.

To the extent an investment manager selling data wishes to take the view that its data sales business is not regulated under the Investment Advisers Act, a threshold question is whether the sales are being made by the SEC-registered investment adviser entity. If so, the presumption will be that the activity is regulated.

Outside the registered entity, it is possible to consider whether the business of a data vendor is beyond the scope of the Investment Advisers Act. In this regard, there is a long history of data vendors approaching SEC to ask exactly that question (am I an investment adviser?). That back-and-forth generated a series of SEC staff interpretive letters over thirty years, which collectively stand for the principle that a data vendor is not an investment adviser so long as:

- (1) The information provided is readily available in its raw state;
- (2) The categories of information presented are not highly selective; and
- (3) The information is not organized or presented in a manner that suggests the purchase, holding or sale of any security or securities.

Given the profusion of data-based businesses today it is somewhat surprising that the last of these letters was issued in the 1990s.⁶³

Looking beyond the SEC, consumer protection and privacy laws regulate data resellers and data brokers (which would cover many data vendors), whether or not the data they sell constitutes a “consumer report” for purposes of the Fair Credit Reporting Act.⁶⁴ And, as with

63. See, e.g., Missouri Innovation Center, Inc., SEC No-Action Letter (Oct. 17, 1995); Media General Financial Services, Inc., SEC No-Action Letter (July 20, 1992); Charles Street Securities, Inc., SEC No-Action Letter (Nov. 28, 1989); Butcher & Singer, SEC No-Action Letter (Jan. 2, 1987).

64. A variety of bills in Congress (proposed by both Democrats and Republicans) address data protection and could direct the FTC to perform a study of the data reseller industry and/or propose regulations. One example is the American Data Dissemination Act, sponsored by Sen. Marco Rubio (R-FL) and introduced Jan. 16, 2019. Press Release, U.S. Senator Marco Rubio (R-FL), Rubio Introduces Privacy Bill to Protect Consumers While Promoting Innovation (Jan. 16, 2019), <https://www.rubio.senate.gov/public/index.cfm/2019/1/rubio-introduces-privacy-bill-to-protect-consumers-while-promoting-innovation>.

There also remains the possibility of administrative agency or executive department rulemaking. The National Telecommunications and Information Administration within the U.S. Commerce Department has had an open statement of regulatory intent in the Federal Register since September 2018. That notice refers to harmonization and advancement of privacy regulation, see Nat'l Telecomms. & Info. Admin., U.S. Dep't of Commerce, Developing the Administration's Approach to Consumer

considering whether the SEC might view provision of data to be regulated “investment advice,” data sold to third parties that could be considered to bear upon a consumer’s credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics or mode of living, depending on its intended use, may be a “consumer report” under the Fair Credit Reporting Act (FCRA), and could thus make the firm a “consumer reporting agency” under the FCRA.⁶⁵ Such status carries significant compliance burdens and the heightened legal risk from consumer claims and regulatory enforcement for violations.

Moving to the states, two current state level registration requirements apply to data brokers, in California and Vermont, which not only require public filings, but enhanced disclosures on data processes and practices.⁶⁶

§ 66:6.3 Data Governance

Related to the core concepts of data protection and data lineage is the broader idea of data governance. At its most basic, data governance is intended to ensure data quality within a firm. The program is dedicated then to the nuts and bolts of maintaining availability of and access to data, data consistency, data mobility, data integrity, and data protection. When dealing with regulated data or a regulated organization holding and using data, as in the case of investment managers, the data governance program will connect to and may overlap with

Privacy, 83 Fed. Reg. 84,600 (Sept. 26, 2018), <https://www.federalregister.gov/documents/2018/09/26/2018-20941/developing-the-administrations-approach-to-consumer-privacy>. However, past efforts by the FTC to regulate data sellers stalled after a series of studies and recommendations sponsored by the agency from 2012 and earlier. *See, e.g.*, Fed. Trade Comm’n, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers (Mar. 2012), <https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers>.

65. *See* 15 U.S.C. § 1681a; *see also* Fed. Trade Comm’n, 40 Years of Experience with the Fair Credit Reporting Act: An FTC Staff Report with Summary of Interpretations (July 2011), <https://www.ftc.gov/sites/default/files/documents/reports/40-years-experience-fair-credit-reporting-act-ftc-staff-report-summary-interpretations/110720fcrareport.pdf>.

66. *See* CAL. CIV. CODE § 1798.99; VT. STAT. ANN. tit. 9, § 2430. For a description of California data broker registration, *see* XAVIER BECERRA, ATT’Y GEN., STATE OF CAL. DEP’T OF JUSTICE, DATA BROKER REGISTRATION, <https://oag.ca.gov/data-broker/register>. For a description of Vermont data broker registration, *see* Press Release, TJ Donovan, Vt. Att’y Gen., Off. of the Vt. Att’y Gen., Attorney General’s Office Issues Guidance on Data Broker Regulations (Dec. 13, 2018), <https://ago.vermont.gov/blog/2018/12/13/attorney-generals-office-issues-guidance-on-data-broker-regulations/>.

the firm's compliance program. As service providers (administrators, transfer agents, custodians) tend to have a central role for investment management businesses, an investment manager's data governance program likely also contemplates significant service provider connectivity and data transfers.

As with any organizational program, data governance requires an "owner," who is ultimately responsible for its implementation. In larger or data-centric organizations, there may be a chief data officer. For investment managers, responsibility most likely will sit with a chief technology officer, chief information security officer, chief operating officer, chief risk officer, or, sometimes, chief privacy officer.

Data governance also can address a firm's view on more philosophical questions, like those of ethics and fairness. It has been common over many years for firms that make heavy use of data to speak of their "data ethics." This is sometimes referred to as embodying the principle that the question for a firm is not whether it can (operationally or legally) put data to a particular use, but whether it *should* (whether doing so is "right"). Data ethics policies are intended to ensure that an organization has a governance framework to answer that question and, in doing so, considers a broad range of factors (for example, legal and contractual requirements, technical capacity, social expectations, reputational considerations, and the like).⁶⁷

Illustrative of the can/should dichotomy is a speech by a former SEC official, who held AI out as a potent tool in developing actionable insights for the agency's examination and enforcement programs. But the official pointedly added:

... algorithms can't then prepare a referral to enforcement. And algorithms certainly cannot bring an enforcement action. The likelihood of possible fraud or misconduct identified based on a machine learning prediction cannot—and *should not*—be the sole basis of an enforcement action (emphasis added).⁶⁸

In other words, AI insights inform enforcement thinking, but when it comes to whether to invoke government authority in a way that

67. As one prominent example in a related field, a condition of the sale of the AI firm DeepMind to Google reportedly involved the establishment of a specialized board to oversee the ethics implications of how DeepMind's AI would be used by Google. See *Whatever Happened to the DeepMind AI Ethics Board Google Promised?*, GUARDIAN (Jan. 26, 2017), <https://www.theguardian.com/technology/2017/jan/26/google-deepmind-ai-ethics-board>. See also *How Big Tech Is Struggling with the Ethics of AI*, FIN. TIMES (Apr. 28, 2019), <https://www.ft.com/content/a3328ce4-60ef-11e9-b285-3acd5d43599e>.

68. Bauguess Speech, *supra* note 47.

can imply or actually alleges wrongdoing (which is what a subpoena or enforcement action does), it is not yet appropriate to give an algorithm the last word.

§ 66:6.4 **Model Governance**

As outlined at section 66:5 above, there have been a number of SEC enforcement actions against firms whose complex investment models allegedly failed them (with the resulting enforcement typically focused on alleged failures to disclose, or even cover up, issues affecting firm clients). The overall impression from the cases is that the SEC expects that a firm should follow what is often referred to as robust “model governance,” being the governance and controls frameworks that wrap around development and use of complex quantitative models. Disclosures of risks, both up front and over time, also should be considered.⁶⁹

From the SEC’s enforcement history, the following might be taken as embodying the agency’s expectations for model governance. In

69. While any disclosure should be thoughtfully tailored to the particular firm’s circumstances, an example of very general risk disclosure relating to the use of AI techniques in trading would be the following:

AI-Based Trading Systems. In line with advances in computing technology and data analytics, there has been an increasing trend towards machine driven and artificially intelligent trading systems, and particularly towards providing such systems with increasing levels of autonomy in trading decisions. There are various risks associated with these systems. Regulators have been increasingly active in considering regulations and market restrictions on algorithmic and other machine assisted trading strategies, including, efforts to require pre-testing of such techniques, to impose automatic volume controls and/or to impose liability for negative or manipulative market impacts of such trading. Such restrictions may also impair the operation of fully autonomous trading systems and technologies, either by design or inadvertently. In addition, such technologies are relatively recent developments and may be subject to one or more undetected errors, defects or security vulnerabilities. Some errors may be discovered only after a product or service has been used by end customers or after substantial operations in the market place. Any exploitable errors or security vulnerabilities discovered after such products are in widespread operation could result in substantial loss of revenues or assets, or material liabilities or sanctions. The potential speed of such trading technology may exacerbate the impact of any such flaws, particularly where such flaws interact with other algorithmic systems and or where various technological or operational limitations may act to impair or prevent the intervention of a human control.

particular, an investment management firm deploying a complex algorithm or quantitative model should:

- (1) Take reasonable steps, both before launch and over time, to determine that the algorithm operates as intended and in compliance with applicable laws and regulations;
- (2) Rigorously test coding and the underlying math (when possible, follow a maker-checker approach, which separates testing from the original coding and design);
- (3) If there are material issues or vulnerabilities related to the algorithm, fix them promptly and make appropriate disclosures;
- (4) If there are material limitations or risks associated with use of the algorithm, disclose them before an issue arises;
- (5) Ensure that data, computing environment, and underlying algorithms are appropriately protected from reasonably anticipated external and internal risks to their security, integrity, confidentiality, and availability;
- (6) Be thoughtful about organizational choices (do not isolate development and oversight of quantitative models outside core management and compliance oversight; do not leave models to inexperienced staff or new hires without taking care to supervise and integrate them with the firm);
- (7) Recognize that applying senior management and compliance oversight to quantitative models requires at least some technical capacity (in other words, control functions need sufficient understanding of the algorithm to ask probing questions);
- (8) Design oversight that draws on diverse functions (portfolio management, trading, technology, security, operations, product support, and legal and compliance all have a role; “new product” protocols at most large firms are cross-functional, but a firm vetting its first quantitative or AI model should not assume existing protocols will be sufficient);
- (9) Recognize that protection of trade secrets and intellectual property are important, but cannot override compliance and fiduciary requirements;⁷⁰ and

70. The SEC enforcement history suggests that the agency believes that intellectual property protections and related tight controls on information at times have frustrated legal, compliance, and even senior management oversight of models.

- (10) Be able to explain the algorithm's core operations, data integrity and controls, and outcomes to the firm's internal and external governance bodies (senior management, compliance and control functions, and regulators). This may involve a significant records retention and management component.

In the FINRA AI 2020 Report, FINRA outlined broadly similar views and confirmed the position that Rule 3110 (FINRA's general rule requiring supervision of firm processes) requires firms to maintain written supervisory procedures (WSPs) and control systems for AI-based tools. Further to that general guidance, FINRA specifically encouraged firms that employ AI-based applications to review and update their model risk management frameworks to address challenges AI models may pose.

Where applicable, FINRA laid out the following as areas for firms to consider as they update their model risk management programs to address AI. Firms should:

- (1) Update model validation processes to account for complexities of a machine learning (ML) model.⁷¹ This includes reviewing the input data (for example, review for potential bias), the algorithms (for example, review for errors), any parameters (for example, verify risk thresholds), and the output (for example, determine explainability of the output);
- (2) Conduct up front as well as ongoing testing, including tests that experiment with different and stressed scenarios (for example, unprecedented market conditions) and new datasets;
- (3) Employ current and new models in parallel and retire current models only after the new ones are thoroughly validated;
- (4) Maintain a detailed inventory of all AI models, along with any assigned risk ratings such that the models can be appropriately monitored and managed based on their risk levels; and
- (5) Develop model performance benchmarks (for example, number of false negatives) and an ongoing monitoring and reporting process to ensure that the models perform as intended, particularly when the models involved are self-training and evolve over time.⁷²

71. Per FINRA, model risk management becomes even more critical for ML models due to their dynamic, self-learning nature.

72. The FINRA report on digital advice, *see infra* note 73, includes an additional lengthy discussion of FINRA's expectations for model governance.

§ 66:6.5 Robo-Advice Considerations

Various regulatory questions have been posited regarding robo-advice, with an emphasis on being sure:

- (1) The client understands the nature and limitations of the service;
- (2) The questionnaire used to interact with the client is appropriate, complete, and thoughtfully designed to gather the right feedback; and
- (3) The algorithm is properly tested and maintained.⁷³

Compliance policies and other internal controls should address each of these questions. Policies and controls also should be prepared to respond to continued extensions of the robo-advice model, for example, as program designers seek to take into account an increasing variety of information on users (for example, through web-scraped social media profiling).

To respond to potential risks of prohibited bias becoming embedded in a robo-advice model, firms also could consider “blind spots” in their data—especially those that may not fully account for needs and interests of customers outside majority or privileged populations. This would be in line with a seminal 2016 FTC report on the intersection of big data and bias, in which the FTC encouraged companies to consider the following.⁷⁴

- (1) Whether datasets are missing information about certain populations and steps that might be taken to address issues of underrepresentation and overrepresentation;
- (2) Whether biases are being incorporated at both the collection and analytics stages of big data’s life cycle;
- (3) Whether correlations identified in the data are, in fact, meaningful; and

73. SEC Staff Issues Guidance Update and Investor Bulletin on Robo-Advisers (Feb. 23, 2017), <https://www.sec.gov/news/pressrelease/2017-52.html>; FINRA, REPORT ON DIGITAL INVESTMENT ADVICE (Mar. 2016), <https://www.finra.org/sites/default/files/digital-investment-advice-report.pdf>.

74. FED. TRADE COMM’N, BIG DATA: TOOL FOR INCLUSION OR EXCLUSION: UNDERSTANDING THE ISSUES (Jan. 2016), <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

- (4) Whether the predictive value of a model fits with fairness considerations.⁷⁵

FINRA's AI 2020 report suggests similar considerations, including that firms should enhance their policies and procedures to include a review of the underlying dataset for any potential built-in biases; firms also should apply specific data filters to assess how the model outputs are affected; leverage proxies in lieu of demographic data; and utilize a diverse group of staffers to review the dataset and test the model outputs.

§ 66:6.6 **Diligence of RegTech Providers**

The promise that technology-enabled compliance solutions will revolutionize the work of regulatory lawyers and compliance officers has, to date, been only partially realized. In some cases, that is because firms that implement RegTech solutions have not always found them to be well suited or properly tailored to a firm's own environment. Developing a solid checklist to evaluate RegTech tools is critical to success. Examples of diligence questions a firm might ask include:

- (1) Whether the tool's designers truly understand the regulatory issue they are solving for;
- (2) Who else has road-tested the tool;
- (3) Whether the intellectual property underlying the tool is in order;
- (4) How the tool will interface with legacy systems at the firm (or often just as important, the firm's service providers);
- (5) Which data sources the tool draws on and whether it will cleanly ingest the firm's data (or, again, service provider data);
- (6) Whether the tool creates new data exposure or security risks;
- (7) What redundancy and business continuity protections are available;
- (8) How much product support and customization is available;
- (9) Whether licensing terms will limit use across geographies, business functions, or with affiliates; and

75. As an example where fairness might be in question, consider that the FTC has found evidence that credit scores for certain groups of people were lowered on the basis of repayment histories of other people with similar preferences in retail stores. *Id.*

- (10) How cooperative the vendor will be in the course of audits (and potentially, inquiries from a firm's regulators into the tool's functions and design).

§ 66:6.7 *Conflict of Interest Considerations and Disclosure—Whose Data Is It?*

All organizations face the question of who owns data derived from customer activity. While in the United States it generally has been presumed that such data belong to the firm, it is this fundamental question that animates much of the data-related criticism of firms like Facebook and Google. There are increasing calls for data subjects to have data rights—and even monetary compensation—for the value that their data brings to the businesses that use them. Indeed, this idea is reflected in part in discussions about the development of principles for a “data economy,”⁷⁶ as well as the CCPA's regulations around discrimination that requires companies to calculate and disclose the value of personal information to the business when providing financial incentives or differential treatment of goods and services depending on whether an individual provides personal information to the business.⁷⁷

In the case of an investment management firm's client-level account data, the view is well established that nearly all data associated with trading decisions by or for a client that might be captured by the firm are owned at least as much by the firm as the client. Similarly, all research undertaken for a client's account, and related intellectual property and know-how generated by the firm when acting as agent for the client, are presumed to belong to the investment manager. While this starting point has not been questioned, some firms have begun to consider these presumptions through the lens of disclosures of conflicts of interest. On that basis, some investment

76. The American Law Institute, jointly with the European Law Institute, is working on a series of papers with the aim of developing legal principles for a “data economy” that could further address these questions of data ownership and monetization. For additional information on the ALI/ELI project, see EUROPEAN L. INST., PRINCIPLES FOR A DATA ECONOMY (WITH THE ALI), <https://www.europeanlawinstitute.eu/projects-publications/current-projects-feasibility-studies-and-other-activities/current-projects/data-economy/>.

77. See CCPA Attorney General Proposed Final Regulations section 999.307 (2020).

management firms have begun to highlight in their client communications their views on data ownership that might be in conflict with a client's interest.⁷⁸

-
78. While any disclosure should be thoughtfully tailored to the particular firm's circumstances, an example of such disclosure (in this case, for a private equity fund manager) illustrates this thinking:

Data Management. We [referring to the investment management firm] receive or obtain various kinds of data and information from portfolio companies, funds and client accounts that we or our affiliates manage or advise (collectively, "Data Holders"), including data and information relating to business operations, trends, budgets, customers and other metrics, some of which is sometimes referred to as "big data." Information obtained from the Data Holders may provide material benefits to us without compensation or other benefit accruing to the Data Holders or their investors [highlighting that the data is being collected by the firm without compensation to clients whose assets fund the investment activity or to the portfolio companies that may be the source of the data]. We generally are free to use data and information from these activities to trade for the benefit of ourselves or another client.

