Introduction

Technological innovation continues to change employer practices, creating new legal challenges for their legal counsel. One significant such development is employers’ growing tendency to use big data to answer their most pressing questions. Once reliant on optimistic revenue predictions or sparse, anecdotal accounts of employee satisfaction, employers and legal counsel now may sift through enormous data sets to answer complex and sophisticated questions about applicants and employees.

Harnessing the power of these massive data sets, or “big data,” allows attorneys to understand historical patterns of legal activity, improve existing employment practices, and even increase the efficiency and efficacy of their own law firms. Armed with these colossal resources, algorithms help employers uncover interactions behind the rise and fall of business revenue, employee productivity, hiring patterns, disciplinary pitfalls, financial risk, legal exposure, and myriad other factors that influence a business plan’s success.

However, using automated machine-based outputs to understand individual human beings’ actions is fraught with risks. Potential

---

* Darrell S. Gay is the partner in charge of the labor and employment group at Arent Fox LLP in New York City where he represents for-profit, non-profit, and government employers.

** Abigail M. Kagan is an associate at Arent Fox LLP in New York City. She represents employers in labor and employment law matters.


exposure to serious legal action looms over every decision based on big data.5 Despite the risks, clear legal rules are largely absent. Lawyers' discussions of big data are rampant with misinformation, likely because few attorneys moonlight as data scientists.

We begin this discussion by defining some terms. Artificial intelligence is the idea that computers can “carry out tasks in a way that we would consider smart,”6 essentially meaning that a computer is capable of being taught to think and understand the world by classifying information as humans do.7 By contrast, “machine learning” is a “an application of [artificial intelligence]” based on the notion that programmers should be able to provide data to computers to allow them to learn on their own.8 Both concepts are important for understanding analysis of “big data.”9

This Article provides an introduction to the fundamentals and breakthroughs of big data and how they implicate employers and employment law. Part I offers a basic understanding of big data and artificial intelligence, focusing on some of the most common sources of big data. Part II discusses how employers can beneficially harness big data's power. Part III explores big data's dangers and pitfalls, particularly arising from machine learning and predictive analytics. Part IV reviews the sparse legal authority on big data. Part V focuses specifically on law firms' potential use of big data and artificial intelligence in firm management.

I. Big Data and Artificial Intelligence

A. Big Data

For employers, “big data” refers to the mass of information created when individual employees engage in trackable activities.10 For example, when employees visit websites at work, employers' computers record which URL was followed, how long the employees remained on the webpage, and what actions they took there. When employees buy lunch at work, barcode scanners note the purchase time and the chosen items. When employees open Microsoft Word documents, software tracks which documents are edited and the time spent on each

---

7. See id.
8. Id.
9. For an explanation of “big data,” see infra Section I.A.
page. When employees file lawsuits, legal search engines can track other complaints filed for the same reason and which causes of action are brought at the same time. The list goes on. When such trackable activities recur multiple times by multiple people, the mass of results accumulates. The aggregate result is known as “big data.”

The following sections describe the process of accumulating big data and some of its uses.

1. Descriptive Analytics

In the “descriptive analytics” stage of data processing, computers gather, organize, tabulate, and depict data. As a very basic example, the calendar software Outlook generates masses of data unintelligible to the average employer. But a computer’s descriptive-analytics output can make sense of that data and translate it into employer useful reports by showing, for example, all dates on which a particular conference room is reserved.

Descriptive analytics can help identify otherwise unseen relationships between two or more data points. It can facilitate employer and attorney decision-making by translating overwhelming and amorphous data sets into actionable intelligence. Such useful information can be made available in real time. Some major sources for employment-focused big data are illustrated below.

2. Public Information

Public information from social media is a prime wellspring for big data. LinkedIn, self-described as “the world’s largest professional network,” holds a massive amount of employment-related data. For all members, LinkedIn identifies how many positions they have held, whether they received accolades in professional or academic spheres, their connections to others, and where their future professional interests lie.

---

14. Id.
Other public sources include Twitter, Snapchat. “Trending hashtags” identify hot topics in real time. City-wide launches of Uber and other ride-sharing services show how city-dwellers travel to and from work and whether they bill these trips to their personal or business accounts within each service’s mobile app. Health insurers market mobile apps that track users’ exercise, permitting insurers to provide discounts to users with healthy habits. All this information contributes to a general understanding of human behavior, from individuals to broader populations.

Some governmental agencies provide good sources of big data. The U.S. Department of Labor’s “enforcement data” website publishes raw datasets of all federal enforcement and compliance actions concluded in the prior decade, searchable by state, company name, amount of back wages owed, and dozens of other categories. Myriad other online sources demonstrate the breadth of big-data resources available on the Internet that employers may potentially harness. OpenFDA, from the Food and Drug Administration, provides transparent information on a variety of issues, including drug product labeling and recall enforcement reports, for example.

3. Company Personnel Information

Most companies permit or require employees to use some sort of company-provided data tracker, such as cell phones, laptops, corporate

---

17. Twitter is an online news and social networking service that “give[s] everyone the power to create and share ideas and information instantly, without barriers.” Our Company, Twitter, https://about.twitter.com/en_us/company.html (last visited July 2, 2018).


23. See Data Enforcement: Data Catalog, U.S. DEP’T OF LAB., https://enforcedata.dol.gov/views/data_catalogs.php (last visited July 2, 2018). For example, to see all compliance actions concluded by the Wage and Hour Division since Fiscal Year 2005, scroll to the bottom of the page and click “WHD Data,” then click the .csv option under “whd_whisard.”

24. See infra Section II.G. for additional discussion on employers’ use of big data from governmental agencies.

credit cards, and digital timekeepers. These computer-based devices generate a trove of big data that employers may then harness to improve hiring, retention, compensation, and talent development.

For example, employers may use Outlook calendar entries to learn the most common times for meetings, which departments are most likely to book conference rooms, and how to maximize office space use. Similarly, payroll logs show the most popular employee vacation times, the extent of annual sick day usage, and the total overtime hours each pay period.

Many employers overlook more old-fashioned information sources, such as exit interviews or employee satisfaction surveys. Such mechanisms that gather narrative information directly from employees permit more insightful analysis of collected data. Studying data points gleaned from payroll figures and other automated sources may identify the probability of two events occurring together, but only employees can explain why those events actually coincide.

4. Employee Grievances

Businesses with formal grievance processes have even richer sources of employee data. With names or other identifiers attached to grievances, employers can track which employees bring the most complaints and whether grievances were considered meritorious. Even anonymous grievances tell companies whether a particular department, supervisor, or policy causes employee problems. “Grievance tracker” websites and apps are on the rise, allowing employers, employees, and unions to file, organize, and respond to grievances online. Using dropdown menus, checkboxes, and other standardized measures, grievance process managers can more easily review and categorize complaints. This process informs employer responses and provides key insights to improve future operations and thereby avoid future grievances.

5. Case/Litigation History

As attorneys know, legal search engines are gold mines for raw data. Westlaw, Bloomberg Law, and other search engines identify


lawsuits filed against particular companies, which causes of action have historically prevailed, how particular judges rule, and even which attorneys most frequently prevail. Such litigation history can allow counsel to advise clients on the likelihood of success before specific judges or what potential damages to expect, while also improving law firms’ internal risk-assessments by informing decisions whether to accept particular cases initially.

B. Artificial Intelligence

Amassing data is only the first step in the process of interpretation. An overwhelming amount of information is tracked, logged, discovered, noted, and revealed every minute of every day. How is such a vast quantity of information converted into useful patterns and lessons?

Historic small-scale employer reviews of employee data, such as surveys or overtime requests, are wholly inadequate for the overwhelming quantities of big data. Employers and businesses that collect big data now recognize that “datasets are ripe for computer-assisted or automated analysis.” Artificial intelligence can fill that need.

While big data afford raw information to form the foundation for descriptive analytics, artificial intelligence provides necessary context and meaning to make sense of otherwise two-dimensional information. Artificial intelligence allows computer systems behind everyday interactions constantly to watch everyday human interactions, identify patterns, and gain understanding by experience. Consider Siri, the ever-present virtual assistant inside every iPhone. When iPhone users pose a question to Siri, “she” listens to commands, parses meanings properly, “thinks” of which words to combine in response, and “understands” which syllables to emphasize.

Similar technology allows employment software to understand that the best employees are those who take short lunch breaks, exhibit risk-averse traits like purchasing family-coverage health insurance, extensively use their Outlook calendars, send frequent emails but not

pestering ones, and clock in and out right on time, but not so early or late as to incur overtime pay. Most importantly, artificially intelligent computers can make these determinations without employers needing to identify relevant factors. In fact, the computer might use factors the employer never considered. Employment software can even produce a periodic ranking of what it has determined to be top employees based on a delicate, weighted, ever-shifting balance of multiple characteristics.

II. Use of Big Data for Employers

As Part I of this Article demonstrated, big data's uses appear limitless, especially when used in concert with artificial intelligence and machine learning. Part II focuses on some of the most beneficial ways employers can harness big data.

A. Selection, Promotion, and Discharge

The clearest applications of digital big data and machine learning in the employment context are in recruitment, training, promotion, and discharge of employees based on office-wide and population-wide statistics.

An advantage of big data is its ability to de-identify and anonymize individual persons. Computerized algorithms also can be designed to ignore certain categories, such as protected classes under Title VII. This combination helps eliminate discretion and unconscious bias, allowing employers to focus on each applicant's fitness for the job in question. The process begins by constructing algorithms to analyze current employees' professional characteristics from timesheets, surveys, number of emails sent per hour, or other available sources. The algorithms can then uncover patterns associated with top performers, and employers can use that information to improve business decisions. For example, if employees who arrive before 9:00 AM and send at least six emails per hour often score the highest on performance evaluations, employers can create an algorithm that highlights

37. See Marr, supra note 6.
40. Walton, supra note 5.
41. See id.
42. See id.
44. See id.
applicants who exhibit those traits or exclude from consideration applicants who do not.

The same system can evaluate current employees. Employers can use information from past evaluations to create an integrated “model employee” profile, and then compare the profile to current employees to identify weak performers. Employers then have two options: (1) train low-performing to improve; or (2) discharge the employee in favor of an applicant who, according to the algorithm, already embodies the ideal worker. Training is likely the better option because it potentially improves the overall workforce and reduces turnover.

One must, however, keep in mind the possible pitfalls within these analyses. For example, employees who arrive early and send many emails certainly appear hard-working, but perhaps they just want to ensure sufficient time to eat breakfast or email about weekend plans before finally turning to work materials an hour later. Employers must remain aware of existing workplace cultures and norms before drawing conclusions based solely on numbers.

B. Leadership Management

Just as employers can use algorithms to spot poor performers, this information can distinguish hidden high achievers. Employers might consider promoting employees who demonstrate characteristics of the ideal performer on the assumption that their workplace attributes demonstrate superior performance. Consistent high performers might be placed into a leadership-development track, priming an ideal subset of current employees for career trajectories that include high-level managerial positions.

C. Policy Review and Revision

Just as big data about employee behavior and performance can be used to evaluate individual employees, it can also be useful to enhance and update company policies. An example of an outdated policy that requires updating is one that requires employees to telephone supervisors if they are running late. Such a policy might create problems for...

46. See id.
47. Id.
48. See Katz, supra note 43.
49. Florentine, supra note 27.
50. Id.
51. Id.
supervisors who want to accommodate modern conveniences, such as email, but feel they must rule by the book.53 Personnel data may show that most late employees email or text supervisors, instead, indicating that policies should be revised.54

Employers can review data on employees’ daily habits to determine whether they are consistent with existing policies and procedures.55 If not, the employer might decide to update policies to reflect current employee practices. This approach may be practical on several levels. Certain policies can cause workplace friction, leading to arguments, grievances, and even lawsuits.56 Grievance analysis may assist revising and updating policies by better reflecting each group’s culture. Rather than retroactively addressing employee complaints as they arise, employers may improve workplace policies and procedures so that potential conflicts can be mitigated or entirely avoided.57

D. Managing Labor Relations

Employers know that grievances impose significant burdens. Aside from the effort needed to address the root of the problem, formal grievance processes remove employees and supervisors from their normal work and disrupt productivity. Union grievances may require arbitration,58 which can become lengthy, costly, and adversarial. Employers can use historical data to review which internal policies or union contract provisions have created the most disputes. They can then proactively address these concerns with the union and attempt to work out an alternative resolution before entering into a new collective-bargaining agreement.

E. Union Avoidance

Employers may be able to use big data to avoid union certification altogether. Union organizing generally arises from employee discontent over current working conditions or bad management.59 As


54. See Survey, supra note 52.


56. See, e.g., id. (discussing claim employees may make in response to employee monitoring); see also, e.g., Lucas, supra note 53 (discussing issues with employer’s attendance policy).

57. See, e.g., Straz, supra note 2 (big data helps employers understand why employees leave).


mentioned above, employee satisfaction surveys and exit interviews provide key insight to real-time employee grievances, concerns, and other factors that can lead employees to seek union representation. This information, if created and maintained in electronic forms and properly analyzed, can allow an employer to make a timely corrective response before any organizing effort begins. To be useful for this purpose, employee satisfaction surveys need to call for detailed answers that allow employers to identify problems with particularity.

F. Benefits

Big data can help employers refine employment benefits to boost productivity. By comparing payroll information and work output, employers can determine whether factors such as taking vacation, or the length of vacations, affects productivity in surrounding work-weeks.60 Employers might also determine whether employees denied paid vacation become less productive.

G. OSHA Issue Prevention

The Occupational Safety and Health Administration (OSHA) recently mandated information from employers that produces a new source of useful big data.61 Since August 10, 2016, OSHA has required many private employers to report to it any fatality, hospitalization, loss of limb, and certain other work-related injuries.62 The regulation does not change any recordkeeping rules. It simply requires employers to submit electronically certain information they are already required to maintain.63

The OSHA regulation may be a harbinger of things to come. While the Trump administration generally opposes new regulations,64 the current rules may increase public access to big data on workplace incidents. The OSHA database provides the necessary foundation for groundbreaking analyses. If other government agencies follow OSHA’s lead, the country may see massive data sets not only about workplace injuries, but also about such things as payroll systems, employee benefit plans, and tax returns. The federal government has already granted access to a significant amount of information through data.gov.65

60. See Florentine, supra note 27.
63. See id. §§ 1902, 1904.
openFDA,66 and the entire Bureau of Labor Statistics website.67 Interested employers need only review those and other government-provided sources to uncover troves of data, waiting to be deployed.

III. The Dangers of Using Big Data

Using descriptive analytics to maximize efficiency, productivity, and employer-employee relations has its limits. It is the next step—machine learning and predictive analytics—that allows employers truly to harness the power of artificial intelligence and advance workplaces into the twenty-first century.

A. Machine Learning and Predictive Analytics

Machine learning is a form of artificial intelligence that occurs when algorithms use data to improve themselves beyond initial human programming.68 If an employer wants to identify employees with particular traits—from punctuality to the ability to manage large projects or do creative problem-solving—machine learning would allow the algorithm to fine-tune the criteria to identify relevant traits to identify more accurately employees exhibiting those traits. Further, machine learning enables employers to make decisions based on more accurate algorithms.

Machine learning can refine an employer’s original instructions to achieve a more functional result. For example, an original one-dimensional instruction may have categorized successful employees as those who use Outlook for scheduling. The algorithm, as self-enhanced, would instead compare Employee B’s calendar, filled with personal appointments and errands, to Employee C’s calendar, which has only one or two meetings each day, but all business related. The original simple algorithm armed with the capacity for descriptive analytics would rank Employee B higher than Employee C, because Employee B scheduled every minute of the day. However, the enhanced algorithm is “smarter” because it can determine that Employee C’s infrequent meetings with more junior employees, labeled “supervision discussion” or “review of training materials,” are more indicative of a high-performer than Employee B’s active days of “doctor’s appointment” and “coffee with Mom.”

A computer capable of machine learning—meaning an artificially intelligent computer—also applies predictive analytics to its ever-growing neural network, enabling it to review information, track that information’s movement, and anticipate an employee’s or employer’s

---

66. See openFDA, supra note 25.
68. Faggella, supra note 39.
next move. An employer uses a form of predictive analytics when, seeing how quickly Employee C rose through the ranks and hoping for similar star workers, tells the computer to look for more job applicants like Employee C. This approach, however, could backfire. Perhaps Employee C is the junior employee requiring daily meetings for “supervision” and “training,” meaning that Employee C requires extensive oversight, a trait that businesses hope to avoid. The employer is counting on the artificially intelligent computer to learn to examine a more holistic picture of Employee C that includes more than just the employee’s scheduling patterns to avoid filling the workforce with needy employees.

Machine learning at its best should enable computers to “learn” how to process data on their own. The introduction of a self-propelled feedback loop tells the computer’s algorithm whether it correctly classified incoming information, and, if not, how to correct the classification in the future—absorbing information from purposeful (human) and incidental sources with little interruption. The computer provides feedback to itself to adjust its own algorithms and methods of processing various data points constantly, until they are no longer data points but data “lines,” “shapes,” “histories,” or “futures.”

B. Impact of Input Source

Machine learning has to start somewhere. While computers equipped with artificial intelligence employ self-propelled feedback loops, humans are necessary for initial coding and data input. However, human oversight can be riddled with mistaken assumptions, and consequent input errors may originate in multiple ways. Some errors might be correctable early, but, if undetected for too long, the computer’s self-education may manipulate input improperly. Awareness of these potential vulnerabilities may be insufficient to mitigate their effects. Too often, the software engineer who regulates the computer’s self-education is not sufficiently related to the people inputting information. For example, if only a small proportion of employees click on an incorrect answer when completing a survey, inaccurate analysis may result. Software that scans employees’ Internet searches may be misdirected if employees spend lunch hours planning upcoming vacations. Occasional unexpected data points entering the software’s algorithm from one or two employees may not make much difference, but

---

69. See Marr, supra note 6.
70. Id.
71. See id.
72. Id.
74. Id.
75. Id. at 680.
errors can multiply. When this happens, the entire algorithm becomes premised on faulty information.

C. What Questions Are Asked and How

A potential problem in the process arises if managers’ questions do not translate well to computer algorithms. An employer may want to understand the circumstances affecting employee productivity, but this question may not be sufficiently specific to produce useful results. The algorithm may initially look for strong correlations among a variety of data sets without any assessment of the functionality of such data. A serious risk occurs if algorithms leave excess gaps between data points. Human brains will naturally try to fill the gaps and may, as a result, draw erroneous conclusions. It is also difficult for employers to assess the legal validity of algorithms used by third-party vendors because they typically seek to protect their algorithms as proprietary information.

D. Question Validity and What Is Not Being Asked

Without sophisticated machine learning, algorithms can use only the information their creators provide. The resulting analysis will be faulty if a crucial factor is missing from the algorithm. Employers should evaluate algorithms’ accuracy and completeness against current information about the employer’s workforce. Testing how algorithms manipulate and analyze familiar sets of big data can help employers identify simple errors. Algorithms can only highlight preexisting connections between data points. They cannot explain why relationships exist or what external factors might affect these relationships. Without understanding origins of relationships among various data, it is impossible to predict what external factors will affect the analysis and how long an algorithm’s results will remain valid.

76. Id. at 652.
77. Id.
78. See id.
82. Marr, supra note 6.
Each employer should consider not only the information it hopes to learn, but also the information it does not care about or even hopes to avoid. In the workplace, this is key for employers covered by statutes prohibiting consideration of legally protected characteristics when making workforce decisions. Keeping in mind that “we often don’t know what we don’t know” can go a long way in preventing algorithms from fabricating baseless connections among unrelated points of information.

E. Disparate Impact Exposure

Perhaps the most common misconception about using artificial intelligence in employment decisions is that if employers enter no demographic information into their computer-based algorithms, the algorithms will provide fair and equal opportunity for all demographics. There is potentially great liability in allowing algorithms to take control without human oversight. Indeed, at its core, disparate impact liability—the idea that a facially-neutral policy still has a disproportionately negative impact on a particular group or groups, when a truly neutral alternative is readily available—is premised on unintentional but palpable discrimination. Employers who use current employees as models for designing algorithms on how to screen applicants must be sure that their existing employee base does not itself reflect biases. Absent such discernment, artificially intelligent algorithms that work through machine learning can magnify existing biases and discrepancies.

Here is an example of how such magnification of bias might occur. Suppose an employer determines that employees with shorter commuting times are more productive, presumably because they get more sleep or have more relaxed commutes. As a result, the employer instructs an algorithm to select for interviews only applicants who live

86. Id. (facially neutral algorithms may unintentionally harm vulnerable classes).
87. “The efficacy of data mining is fundamentally dependent on the quality of the data from which it attempts to draw useful lessons. If these data capture the prejudicial or biased behavior of prior decision makers, data mining will learn from the bad example that these decisions set. If the data fail to serve as a good sample of a protected group, data mining will draw faulty lessons that could serve as a discriminatory basis for future decision making.” Solon Barocas & Andrew D. Selbst, Big Data’s Disparate Impact, 104 Calif. L. Rev. 671, 697 n.113 (2016).
in two nearby zip codes. The unbiased computer dutifully narrows the employer’s pool of applicants to those two zip codes, which are home to predominantly white, affluent college graduates. Title VII lawsuits ensue. Similarly:

If the training phase for a big data algorithm happened to identify a greater pattern of absences for a group of people with disabilities, it might cluster the relevant people together to create a “high absenteeism risk” profile. The profile need not be tagged as “disability”—rather it might appear to be based on some group of financial, consumer, or social media behaviors.

If a rejected disabled applicant fitting this profile learned of this program, a failure-to-hire lawsuit could follow.

An employer must ensure that the composition of its workforce reflects the qualified population from which that workforce draws. Without advanced machine learning, computers and algorithms do not easily make such nuanced judgments. An algorithm can learn the population makeup of various protected groups, and it can learn the traits that an employer seeks in new employees. However, it cannot adequately balance those potentially competing factors. To defend against allegations of hiring discrimination, employers generally must be able to explain why one employee was hired rather than another qualified applicant who is a member of a protected class. When employers rely too heavily on algorithms that, even innocently, do not receive proper “instruction” and oversight, there is potential legal exposure.

The possibility of liability raises several concerns for employers. For example, algorithm development is usually outsourced to a third party, but liability for any discriminatory effects of the algorithm’s use falls on the employer. The algorithm developer’s liability, if any,

---

89. For an example of software that calculates which postal codes are within a predefined travel time from a certain location, see Find Out Which Postcodes Are Within 30 Minutes Travel Time, TRAVELTIME PLATFORM, https://tools.traveltimeplatform.com/#/postcodes-within-a-radius/request-sample (last visited July 2, 2018).


92. But see Barocas & Selbst, supra note 87, at 700 (to be liable “under current doctrine, the employer would likely both have to know that this is the specific failure mechanism of the model and choose it based on this fact”).


95. King & Mrkonich, supra note 79, at 583.

96. Statutory liability under Title VII is placed on employers—not third party vendors. Id. at 557; see also Martha Chamallas, Two Very Different Stories: Vicarious Liability Under Tort and Title VII Law, 75 OHIO ST. L.J. 1315, 1319 (2014); cf. EEOC v.
is unclear.97 Further, the “black-box” nature of algorithmic decision-making means that employers may be penalized for disparate impacts they could not avoid because they were unaware that such outcomes were likely.98 If employers do not know the assumptions underlying an algorithm’s decision-making, they cannot address any problem with the algorithm’s output. Machine learning further distances results from employer oversight because its fundamental purpose is to operate without employer participation.99 The enormous volume of data means that there is no practical way for employers (or third parties) to spot or correct misdirected algorithms.100

Employers should be alert for other areas of law that change in response to the growing use of big data. For example, one current concern is the ability to apply predictive analytics to data from employer-sponsored health insurance plans to determine whether certain categories of employees are likely to develop illnesses or incur claims.101 In response, some have advocated amending the Americans with Disabilities Act (ADA) to ensure continuing shelter for persons regarded as disabled, or even as potentially disabled.102 It may be only a matter of time before lawmakers recognize such risks of big data and update the ADA and other laws.

IV. Current and Future Legal Implications

The lack of relevant legal authority compounds the legal risks arising from employer use of big data analytics. With no comprehensive case law or policy directives available, employers and their counsel can find only piecemeal legal guidance.

Kronos Inc., 620 F.3d 287, 296 (3d Cir. 2010) (administrative subpoena enforced against non-party provider of assessment used by employer in disparate impact case).

97. Peter Siegelman, Contributory Disparate Impacts in Employment Discrimination Law, 49 WM. & MARY L. REV. 515, 521 (2007) (“The problem of two-party causality in disparate impact suits has not been widely recognized or adequately addressed by either courts or scholars.”). A possible solution for employers could be to seek prior indemnification from the algorithm developer. However, this would likely increase development costs. See King & Mrkonich, supra note 79, at 583.

98. Barocas & Selbst, supra note 87, at 674 (“Discrimination may be an artifact of the data mining process itself, rather than a result of programmers assigning certain factors inappropriate weight.”).

99. T.C., supra note 80.


A. Governmental Agency Oversight and Directives

Several federal authorities have begun to recognize the benefits and risks of big data. A 2016 Federal Trade Commission (FTC) report highlighted issues relevant to employer use of employee and applicant credit reports.103 In that report, the FTC promised to continue to monitor how big data could potentially violate the Fair Credit Reporting Act and the Equal Credit Opportunity Act.104 Indeed, the FTC has urged data brokers, who supply employers with reports on applicants' digital lives, to give applicants access to their own reports.105 The Obama White House also published several reports warning of big data's impact on discrimination, noting that algorithms have the potential to “encod[e] discrimination in automated decisions.”106 In October 2016, the Equal Employment Opportunity Commission (EEOC) hosted a special meeting titled “Big Data in the Workplace: Examining Implications for Equal Employment Opportunity Law.”107

While government agencies have recognized multifaceted concerns about big data, they have provided little guidance in statutes or regulations. Instead, employers and their attorneys must rely on case law from old-fashioned employment disputes and hope their predictions prove accurate.108

B. Existing Authority

One resource that highlights the disparity between legal authority and the technological advances of big data is the Uniform Guidelines for Employee Selection Procedures.109 Published in 1978, the Guidelines are the joint product of the EEOC, the Civil Service Commission, the Department of Labor, and the Department of Justice, and “are designed to assist employers . . . to comply with requirements of Federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin.”110 For forty years, various federal agencies have applied the Guidelines to determine

---

104. Id. at 33.
106. Seizing Opportunities, supra note 85, at 45; see also Report on Algorithmic Systems, supra note 81, at 15.
108. See Kroll et al., supra note 73, at 698 (“The United States has a long history of dealing with these ambiguities [in legislation and guidance] through after-the-fact and retroactive oversight by the courts.”).
110. Id. § 1607.1(B).
whether employee-selection procedures comply with various anti-discrimination provisions.\textsuperscript{111} If an agency finds a selection procedure discriminatory, to mitigate its liability the employer must validate the procedure in accord with the Guidelines or other applicable authority.\textsuperscript{112}

However, there is a disconnect between the basic premise of the Guidelines and contemporary use of computer-based algorithms. The Guidelines—published before the existence of the internet—assume employers can explain the basis of their selection procedures.\textsuperscript{113} With computer-based decision-making, employers simply are not able to delineate each step of the process. As a result, validating selection procedures against the Guidelines is at best complicated and, at worst, impossible.\textsuperscript{114} Employers can present a rationale they hope will suffice, but black-box algorithms can provide only results of internal processes, not a map of how results were reached.\textsuperscript{115}

Case law is equally lacking. In June 2018, a Westlaw Boolean search for “adv: ‘big data’ and ‘disparate impact’” returned no employment-related cases. This result suggests either no case law addresses these topics or that existing cases have used different terminology. If different terms are used in litigation to describe relevant technological processes, the relevant case law will be harder to research, and it will take longer to build coherent guidance. Before employers, judges, and scholars coalesce upon the correct terminology, courts and enforcement agencies may see increased disparate-impact proceedings brought under existing anti-discrimination laws and policies.\textsuperscript{116}

V. Use of Big Data and Artificial Intelligence by Law Firms

Beyond the legal and policy concerns of applying big data and machine learning to employment decisions, big data has extraordinary applications in business more generally. For example, law firms increasingly rely on big data and artificial intelligence.\textsuperscript{117}

Firms may use big data and predictive analytics to determine the potential actions, reactions, and legal needs of current and potential clients. Predictive analytics can triangulate data from court e-filing systems, mergers, and other sources to alert attorneys to client

\begin{itemize}
  \item \textsuperscript{111} Id.
  \item \textsuperscript{112} See id. § 1607.5.
  \item \textsuperscript{113} Id. § 1607.4(A) (“Each user should maintain and have available for inspection records or other information which will disclose the impact which its tests and other selection procedures have upon employment opportunities of persons by identifiable race, sex, or ethnic group . . . in order to determine compliance with these guidelines.”).
  \item \textsuperscript{114} See Jay-Anne B. Casuga, Big Data’s Big Challenges for Contractor Compliance, BLOOMBERG L.: NEWS (Dec. 23, 2016), https://www.bna.com/big-datas-big-n73014449039.
  \item \textsuperscript{115} Id.
  \item \textsuperscript{116} See King & Mrkonich, supra note 79, at 563.
\end{itemize}

Attorneys may also use big-data analytics to learn more about their own performance so they can serve and understand clients more efficiently. Some law firms retain entire teams of analytics specialists to mine general and firm-specific data to maximize efficiency.\footnote{119}{See Michael McDonald, \textit{Data Analytics for Lawyers}, ABOVE L. (Feb. 28, 2017, 5:01 PM), https://abovethelaw.com/2017/02/data-analytics-for-lawyers/?rf=1.} Attorneys can use big data to evaluate their success in particular types of cases to understand where best to invest their time and what areas are best left to others.\footnote{120}{See, e.g., Ron Friedmann, \textit{Is 2017 the Year of Legal Data Analytics? Jackson Lewis’ New Team}, PRESM LEGAL, https://prismlegal.com/2017-year-legal-data-analytics-jackson-lewis-new-team (last visited July 2, 2018).} Big data may also be used to analyze how clients receive and use information the firm distributes.\footnote{121}{See Pfeifer, \textit{supra} note 118.} Analytics companies can, for example, tell a firm which recipients have opened the firm’s e-mails and which unsubscribed.\footnote{122}{See David Ricketts, \textit{10 Ways for Law Firms to Use Big Data}, LINKEDIN: PULSE (Oct. 13, 2015), https://www.linkedin.com/pulse/10-ways-law-firms-use-big-data-david-ricketts.} Artificial intelligence and machine learning are playing an ever-growing role in e-discovery, particularly in document review. Once known as the drudgery assigned to junior associates, document review is increasingly accomplished by computers.\footnote{123}{Id.}

\section*{Conclusion}

Big data and computer-based analytics have great potential to optimize workforces, provide equal employment opportunities, and streamline general business and employment practices. However, in applying these new tools, employers and their counsel must exercise constant vigilance to ensure compliance with current and evolving legal authority. Even as lawyers grapple with new legal issues facing employer-clients, attorneys themselves are benefitting from new ways of practicing law. With proper understanding of big data and analytical tools, employers and their attorneys can use big data and analytics to thrive in this burgeoning world of new technologies.

\footnote{121}{See Pfeifer, \textit{supra} note 118.}
\footnote{123}{Id.}