What Lawyers Need to Know About Artificial Intelligence and Cybersecurity

Presented by:

Corey West, Sr. Manager, Litigation Analytics for Summit, LLC
Eulonda Skyles, Esq., Assistant General Counsel for Privacy, Capital One
Barbara L. Johnson, Esq., Of Counsel, Potter & Murdock, P.C.

Sponsored by:

Summit, LLC
Potter & Murdock, P.C.
Minority In-House Counsel Association (MIHCA)
National Association of Minority and Women Owned Law Firms (NAMWOLF)

Ms. Johnson thanks Andy Knauss for his invaluable assistance in the preparation of this paper.

I. Introduction

Technological advancements are having a tremendous impact on the practice of law, particularly employment law. As ownership of smartphones and tablets has become commonplace over the past decade, the use of software, search engines, and other technology in the legal profession has transformed everything from how cases are litigated to how associates are hired. These changes are also having a substantial impact on how legal services are delivered, with technology "opening up" the law to the general public in ways previously unimaginable. Three areas in particular have had the most pronounced effect: big data, artificial intelligence, and cybersecurity. This paper will discuss how employment lawyers are relying on innovations in these areas to manage their clients, and firms, more effectively and efficiently.

II. What Is Big Data?

Employers and lawyers alike have gained the ability to quickly sort through vast, complex collections of data, using algorithms to convert hundreds of facts and figures into thousands of data points.[1] By analyzing this data, lawyers can better understand patterns of legal activity, improve existing firm policies to reveal cost savings, identify risks associated with representing certain clients, determine how particular projects or matters should be priced, and narrowly tailor solutions for specific legal needs.

A. Big Data Defined

"Big data" is the aggregate information generated when individuals engage in trackable online activities.[2] Big data can come from many different sources: traditional desktop computers, mobile transactions and payments, email, interconnected computer devices and sensors embedded in everyday objects (known as the "internet of things"), and countless other digital clicks and interactions.[3] Dramatic improvements in computer processing power and data storage are enabling the analysis of data quantities that, a decade ago, would have been infeasible if not impossible. In addition, advanced algorithms can piece these seemingly unconnected bits of data together, providing valuable insights into individual users and entire markets alike.

There are three types of analytic options with which to examine big data. First, descriptive analytics uses data mining and aggregation techniques to summarize relationships between two or more data points ("what has happened?"). Second, predictive analytics uses statistical models to predict future events or behavior ("what will happen?"). Third, prescriptive analytics seeks to explain why certain outcomes will occur and identify ideal paths or solutions to pursue ("what should be done?").[4] Businesses routinely use the first two types of analytics already, while prescriptive analytics is a relatively new field. Relying on a combination of algorithms, machine learning, and computational modeling procedures, it is likely to have a substantial impact on how

[1] David J. Walton, *Big Data – What It Is And Why You Should Care*, INSIDECOUNSEL, at 1 (Feb. 14, 2014).
[2] *See id.*
[3] Marcus Evans & Mike Rebeiro, *Big Data: Protecting Rights and Extracting Value*, PRACTICAL LAW MAGAZINE, at 1 (Jan. / Feb. 2015).
[4] *See generally* HALO, *Descriptive, Predictive, and Prescriptive Analytics Explained*, https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/ (last visited Feb. 13, 2018).

businesses make decisions and create value (e.g., identifying tasks that must be completed to optimize the rollout of a new product or service across multiple jurisdictions).

B. Growth of Big Data

As data sets have become fuller in content and detail, the emergence of cloud-based storage options such as Dropbox and Google Drive has sharply driven down the cost of maintaining Electronically Stored Information (ESI). With this increased access and affordability, big data has become big business, and companies are scrambling to get their hands on as much of it as possible. On average, they are spending $8 million each year on big data-related initiatives, and overall spending now exceeds $100 billion.[5] For example, employers are using big data to measure employee productivity—keeping track of which websites an employee visits on her work computer and how long she spends working on her projects.[6] In the context of practicing law, software can track when and where a lawsuit is filed, the number of complaints filed against the same party or parties, and the causes of action alleged in these complaints.

However, as trackable activities accumulate over time, the data sets become so large and unwieldy that their size exceeds the ability of standard software tools and database management methods to capture, store, and process within reasonable timeframes. Only ten percent of big data is "structured" in database tables; the majority is unstructured, coming from messier sources such as emails, audio and video files, website browsing, and mobile calls.[7] Indeed, data collection trends suggest more and more data is being collected from mobile environments (e.g., mobile-only apps), providing information about a user's location, website purchases, and social interactions.

This lack of structure is further compounded by the prevalence of "dark data," information collected and retained in the course of regular business activities but not used for decision-making purposes.[8] The costs required to analyze dark data would be prohibitively expensive for most companies; thus, more data is being obtained than can realistically be processed, with this underutilization resulting in waste and missed opportunities. To successfully interpret (and monetize) this raw information, businesses must rely not only on computer algorithms but human data experts to review the four Vs: volume ("how much data?"), variety ("what kind of data?"), velocity ("how fast can the data be analyzed?"), and veracity ("is the data reliable?").[9]

C. Sources of Big Data

i. External Sources

Social media sites provide businesses with a cornucopia of easily accessible information, where text-mining programs can readily spot emerging patterns and trends. Through Twitter and

[5] Will Landecker & John Rake, *Big Data & The Law* (2014), *available at* https://harrang.com/wp-content/uploads/2014/07/Big-Data-and-the-Law.pdf.
[6] Darrell S. Gay & Abigail M. Lowin, *Big Data in Employment Law: What Employers and Legal Counsel Need to Know*, ABA LEL CONFERENCE, at 1 (Nov. 2017).
[7] Will Landecker & John Rake, *supra* note 5.
[8] Sony Shetty, *How to Tackle Dark Data*, GARTNER (Sept. 28, 2017), *available at* https://www.gartner.com/smarterwithgartner/how-to-tackle-dark-data/.
[9] Jobst Elster, *Hype, Reality, Myth or Legend*, LEGAL MANAGEMENT, at 36 (Oct. / Nov. 2013).

Facebook, observers can track real-time fluctuations in public opinion just as easily as brokers can track the daily upticks and downswings of the Dow. LinkedIn in particular has become *the* source to gather employment-related data: based on an individual's profile, employers can learn about her job experience, academic background, accolades and awards she has received, and composition of her professional network.[10] This information can be further extrapolated to predict where her career ambitions lie, which is highly relevant for recruitment purposes. Thus, when an employer needs to scout externally for talent, it will often run a series of LinkedIn searches (perhaps focusing on competitors' employees) and use the results to compile a list of promising candidates.

Beyond social media, businesses have started relying on data from both private sources (with the data proprietary or licensed) and publicly "open" sources (with the data freely available to use and republish without intellectual property restrictions). Open data sets are usually licensed on terms similar to open source software agreements, which (i) permit access to source code and (ii) contain a "right to redistribute," along with an obligation to make any subsequent improvements available on identical terms.[11] In terms of potential liability, businesses should be wary when licensing open data sets, since the aforementioned agreement terms often do not guarantee the data's reliability or non-infringing nature.

Several organizations, including government agencies, have shown a willingness to share data they have collected with the general public as well. While the Trump administration has firmly committed itself to deregulation, current rules may be utilized in ways to promote greater public access to agency data, such as data relating to food recalls that is available on the Food and Drug Administration's OpenFDA website.[12] Similarly, the Occupational Safety and Health Administration (OSHA) now requires many private employers to report any employee fatality, hospitalization, loss of limb, or other significant work-related injury to the agency.[13]

ii. Internal Sources

Companies also collect copious amounts of personal information about their own employees and job applicants. Many employees are required to use company-provided cell phones, laptops, and credit cards, which passively record their everyday work-related activities.[14] Digital calendars contain information about project deadlines and departmental meetings, and payroll logs track how many vacation or sick days an employee takes each year.

This monitoring has raised concerns regarding the extent to which employers can "spy" on employees and whether employees have any reasonable expectations of privacy while on the job. In certain circumstances, there are statutory limits: federal law protects the right of employees to engage in union-related speech over work email systems, though employers may access the

[10] Darrell S. Gay & Abigail M. Lowin, *supra* note 6, at 2.
[11] *See* OPEN DATA HANDBOOK, *What is Open Data?*, http://opendatahandbook.org/guide/en/what-is-open-data/ (last visited Feb. 13, 2018) (emphasizing the importance of open access to promote the "interoperability" or intermixing of data sets, enabling the construction of larger and more complex systems).
[12] *See generally* OPENFDA, https://open.fda.gov/ (last visited Feb. 13, 2018).
[13] *See* 29 C.F.R. § 1904.1. The regulation does not change any record-keeping rules; rather, employers must electronically file (or "e-file") certain information they are already required to maintain. Also, there is a partial exemption for employers of ten or fewer employees, and certain industry groups are exempt as well.
[14] Darrell S. Gay & Abigail M. Lowin, *supra* note 6, at 3.

underlying data generated by such emails for business-related purposes.[15] Also, while studying data points generated from automated sources is well and good, employers should not forget to probe employees for information directly via satisfaction surveys, exit interviews, and formal grievance procedures. While collecting and evaluating such information can be more time-intensive due to the greater range of individual responses, it may be more comprehensive and enlightening than data generated incidentally via software and electronic devices.

D. Gradual Acceptance of Big Data Use by Lawyers

All lawyers, especially those who represent employers, should have a basic understanding of how big data affects their clients and practices. Lawyers have traditionally been slow to buy into the hype surrounding new technology. However, over the past few years, large employment law firms such as Drinker Biddle and Littler Mendelson have added chief data scientist (CDS) roles to their respective leadership teams.[16] These moves could hint at a growing acceptance of data analytics within the legal industry, which was severely affected by the 2008 recession and has yet to fully recover. As the financial crisis exposed latent weaknesses in old business models, companies began leaning heavily on technology to streamline their operations, to great success, forcing many lawyers to reevaluate their hesitation towards innovation or risk being left behind.

Client expectations have accelerated this attitude shift among lawyers. Companies, especially those in the tech world, take advantage of new technology all the time to improve their job performance and business operations, and they expect their outside counsel to do the same. Another factor is competition, which law firms face from (i) other firms that have successfully incorporated analytics and automation into their service offerings, (ii) "lawtech" start-ups, smaller firms (usually founded by ex-lawyers) that are disrupting established business models through their aggressive use of technology, and (iii) accounting firms that have begun to offer legal services.[17] These factors are intertwined with each other: as big data plays an increasingly influential role in the business world, firms without designated CDSs or meaningful investment in digital infrastructure may find themselves at a competitive disadvantage with firms that have adapted to the changing landscape. There certainly has been pushback to these developments as well, particularly among larger firms in the Am Law 200 which can be very entrenched in their "old-school" (and historically effective) practices.[18] This resistance could open the door to smaller firms to become big data innovators, as well as serve as a cost and quality differentiator for firms.

F. Big Data as Part of Legal Practice: Benefits and Drawbacks

[15] *Purple Communications, Inc., and Communications Workers of America, AFL-CIO*, 361 N.L.R.B. No. 126 (Dec. 11, 2014) (finding that employees have a presumptive statutory right to use an employer's email system to engage in protected speech during nonworking time, though some employer restrictions on email access may be proper).
[16] Ian Lopez, *Analytics in Practice: How Data Scientists Can Change Legal*, LEGALTECH NEWS at 1-2 (Feb. 11, 2016), https://www.legaltechnews.com/printerfriendly/id=1202749465738. In addition, Littler Mendelson has developed subscription-based cloud software that any employer can use. The software automatically generates a report analyzing whether the employer should treat a particular person as an employee or an independent contractor.
[17] Jane Croft, *Artificial intelligence disrupting the business of law*, FINANCIAL TIMES (Oct. 5, 2016), https://www.ft.com/content/5d96dd72-83eb-11e6-8897-2359a58ac7a5
[18] *See* Ian Lopez, *supra* note 16, at 4.

Big data generated from legal research can serve as a potent analytical and leveraging tool not just in client counseling matters but also in general firm management:

- WestlawNext, LexisNexis, and other legal search engines abound with raw data, and firms can customize their subscriptions to these search engine based on actual usage (i.e., the average number of hours each lawyer spends researching cases, as well as the areas of law being focusing on during this research).
- Data mining tools such as Aderant Spotlight[19] and LegalVIEW[20] are equipped to aggregate big data sets, revealing pertinent case information such as the number of lawsuits filed against an employer, what causes of action were brought, and how the lawsuits were resolved.
- Predictive analytics can triangulate data from court e-filing systems, social media sites, and other sources to alert lawyers to client compliance and litigation needs before the clients are aware of these needs themselves.
- Corporate law departments can use tools like the TyMetrix 360º mobile app[21] to analyze legal invoices, figure out which factors (practice area, location, etc.) have the greatest effect on billing rates, and subsequently apply this analysis to negotiate more favorable rates with clients.
- Firms can decide, from a risk-management standpoint, whether to take on certain cases at all, weighing the likely outcome of a cause of action in a particular jurisdiction against the costs of litigation and red flags raised by a potential client's experiences with prior counsel (e.g., a history of malpractice suits). These considerations will impact the number of cases that end up going to trial, especially in situations where analytics enables both sides to agree on a case's value and thereby gravitate towards a sensible settlement range.
- Lawyers can even use big data to analyze their own success rates with different types of cases, allowing them to determine how best to allocate their limited time and resources and which projects to delegate to associates or specialists.

While analytics may inform some firm decision making, a dearth of talented CDSs who are also well-versed in law means that some firms are struggling to interpret data they have collected. The natural tension between technological efficiency and the billable-hour model has also discouraged some lawyers from fully embracing and experimenting with big data, to the chagrin of clients eager to reduce their legal fees.[22] A greater reliance on big data likely will result in lawyers needing fewer hours to complete certain tasks and represent clients in certain matters, but refusal to evolve with the times could motivate clients to shop around for cost-effective alternatives, having more of their work performed either in-house (with the aid of legal software programs) or by cheaper specialty firms. Ultimately, this "either-or" situation ignores the reality

[19] *See generally* ADERANT, https://www.aderant.com/solutions-business-intelligence-spotlight/ (last visited Feb. 13, 2018).
[20] *See generally* WOLTERS KLUWER, http://www.wkelmsolutions.com/products/legalview-analytics-offerings (last visited Feb. 13, 2018).
[21] *See generally* WOLTERS KLUWER, http://www.wkelmsolutions.com/products/T360 (last visited Feb. 13, 2018).
[22] *See* VALIDATUM, *The perfect pricing storm; artificial intelligence and hourly billing*, https://www.validatum.com/articles/the-perfect-pricing-storm-artificial-intelligence-and-hourly-billing (last visited Feb. 13, 2018).

that big data is here to stay, and that analytics has a tremendous capacity not just to improve efficiency but to create new opportunities through this efficiency.

III. Advent of Artificial Intelligence

To interpret the huge amounts of data they are accumulating, employers are turning to artificial intelligence (AI). Also known as cognitive computing, AI is used primarily for predictive analytics—identifying patterns within seemingly random collections of data that hint at future trends.[23] Far removed from its sensational depictions in science fiction as self-aware robots bent on the destruction of human society, AI is being used as just another business tool. However, as AI becomes more human-like in its ability to reason, and given AI's superhuman productivity and speed, some lawyers—particularly young lawyers and law students—fear AI will destroy not their lives but their job prospects, by rendering entry-level and legal research positions obsolete.

A. Effect of Machine Learning on Employment Practices

One core processing technique is "machine learning": a computer is provided with data and "learns" how to process it rather than being explicitly programmed to do so.[24] AI uses algorithms to categorize information in certain ways, while machine learning introduces feedback loops that inform the algorithms whether they have correctly categorized the information, and if not, how to remedy this incorrect categorization in the future. Relying on feedback loops, the computer can continually course-correct, adjusting how its algorithms are processing data points.

In the context of employment, computer algorithms involved in the hiring process can be taught to ignore categories of potentially prejudicial information, such as a job applicant's gender, ethnicity, age, and other (self-reported) demographics. This selective data processing could help guarantee that hiring managers focus solely on applicant experience and other qualifications not as prone to discrimination or unconscious bias. The algorithms also may be taught to identify patterns associated with top employees, allowing employers to create a "model employee" profile and use it to detect (i) promising talent who can be pipelined into leadership-development roles and (ii) weak performers who can either be terminated, receive remedial training, or be reassigned to positions for which they are a better fit within the organization.

Machine learning can be used to evaluate not just employees but entire companies; for example, whether existing policies are serving their intended purposes or are negatively affecting employee relations.[25] Certain policies may be causing unforeseen friction in the workplace, leading to confrontations, low morale, and even lawsuits. If a policy is found to have counterproductive side-effects, the employer might use this discovery as an opportunity to update the policy to be more in-line with the daily habits and practices of employees.[26] Rather than retroactively addressing grievances, employers could design new conflict resolution procedures

[23] Julie Sobowale, *How artificial intelligence is transforming the legal profession*, ABA JOURNAL (Apr. 2016), http://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession.
[24] Bernard Marr, *What is the Difference Between Artificial Intelligence and Machine Learning?*, FORBES (Dec. 6, 2016), https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-andmachine-learning/#196ab2982742.
[25] Darrell S. Gay & Abigail M. Lowin, *supra* note 6, at 6.
[26] *Id.*

that mitigate or even eliminate future conflicts. Likewise, by comparing payroll information with work-product output, employers could fine-tune salaries and benefits to boost productivity.

B. Do e-Lawyers Dream of Electric Juries?

AI and machine learning have the potential to fundamentally transform how legal services are created and provided. On an administrative level, AI can monitor compliance with billing guidelines, classify and analyze time entries, and look for inefficiencies in law firm staffing and workflow.[27] Even more interestingly, some AI programs are learning how to think like lawyers. One example, perhaps a harbinger of things to come, is ROSS Intelligence, billed as the world's first AI attorney.[28] The platform relies on the natural language processing (NLP) capabilities of IBM's Watson, an AI supercomputer that gained widespread popularity for its stint on TV trivia show *Jeopardy*, where it managed to defeat past (human) champions.[29]

ROSS was designed to read and understand language, postulate hypotheses when asked questions, compile research, and generate appropriate responses. ROSS's programmers are working with 20 law firms to simulate workflow examples and test results with human feedback. ROSS has also partnered with BakerHostetler to assist the firm with its bankruptcy practice, performing a "legal research" function as it sifts through thousands of documents.[30] There are plans for ROSS to expand into other areas of law as well, such as labor and employment.

C. Will Expert Systems Replace Associates?

Many legal services use a combination of general knowledge (case law, regulatory decisions, official guidance) and "personal" or specialized knowledge, such as how to prepare the most convincing argument for a party's side, taking into account the unique variables presented by the case. Lawyers use this knowledge (data) to predict outcomes, yet the vast amount of information available means that these predictions are usually based on small data sets. To address this limitation, apps called "expert systems" have emerged, computer systems that attempt to imitate the decision-making of lawyers and other humans with expert knowledge.[31]

Expert systems are software programs designed to solve complex problems through reasoning rather than through conventional procedural code. With respect to the legal profession, expert systems can essentially automate a lawyer's expertise—that is, provide accurate and

---

[27] Peter Krakaur, *Artificial Intelligence (AI) in Law Departments: Opportunities* (Oct. 5, 2016), *available at* https://www.linkedin.com/pulse/artificial-intelligence-ai-law-departments-peter-krakaur/.

[28] John Mannes, *ROSS Intelligence lands $8.7M Series A to speed up legal research with AI*, TECHCRUNCH (Oct. 11, 2017), https://techcrunch.com/2017/10/11/ross-intelligence-lands-8-7m-series-a-to-speed-up-legal-research-with-ai/.

[29] THE ATLANTIC, *Watson Takes the Stand*, at 4 (Jan. 16, 2017), *available at* http://www.theatlantic.com/sponsored/ibm-transformation-of-business/Watson-takes-the-stand/283/.

[30] Karen Turner, *Meet 'Ross,' the newly hired legal robot*, THE WASHINGTON POST (May 16, 2016), *available at* https://www.washingtonpost.com/news/innovations/wp/2016/05/16/meet-ross-the-newly-hired-legal-robot/?utm_term=.84bf9270e941.

[31] Greg Wildisen, *Is Artificial Intelligence the Key to Unlocking Innovation in Your Law Firm?*, LEGALWEEK (Nov. 12, 2015), http://www.legalweek.com/sites/legalweek/2015/11/12/is-artificial-intelligence-the-key-to-unlocking-innovation-in-your-law-firm/?slreturn=20180106164640.

immediate answers for specific legal questions.[32] While expert systems cannot identically replicate human judgment and discretion (yet), they can easily perform tasks historically done by associates. These systems can be customized around particular firm practice areas, enhancing the value and relevance of the content analysis. In addition, services such as LegalZoom, LawDepot, and HotDocs can take user-submitted information and automatically draft basic legal documents (contracts, wills, promissory notes, patent applications) in a matter of minutes, as opposed to hours (if done by a human).[33] Even computer-generated drafts of complex legal documents such as briefs and memos—which require specialized and more sophisticated information—can serve as helpful models, guiding associates as they revise and polish the drafts into their final versions.

Legal work that involves reviewing wide swathes of raw data and isolating bits of useful information lends itself to being automated.[34] A growing number of tools that perform this work are available to lawyers on the market,[35] including Lex Machina (uses data derived from thousands of patent cases to predict the outcomes of infringement suits),[36] LexPredict (can predict the outcomes of Supreme Court cases, and the votes of individual Justices, at rates equivalent to human experts),[37] Kira Systems (facilitates contract review work by automatically highlighting and extracting key language, in any file format),[38] and ThoughtRiver (scans and interprets language in written contracts for risk-assessment purposes).[39] With these tools at their disposal, associates can spend less time on mundane busywork and more time as "project managers" on higher-value matters. Going forward, this technology-borne shift could have a huge impact on firm hiring practices, the training junior lawyers receive upon joining a firm, and even the job descriptions of associates, with perhaps a greater emphasis on coding/programming and data analysis.[40]

D. Cost Savings through e-Discovery and "Smart" Searches

AI also features prominently in the fields of regulatory compliance and electronic discovery (or "e-discovery"). Document review is by far the most expensive and time-consuming part of discovery, with multiple attorneys (again, usually associates) tasked with wading through millions of documents—which can span many years and several jurisdictions—to determine which

---

[32] Lorelei Laird, *Expert systems turn legal expertise into digitized decision-making*, ABA JOURNAL (Mar. 17, 2016), http://www.abajournal.com/news/article/expert_systems_turn_legal_expertise_into_digitized_decision_making.

[33] Ben Allgrove & Yoon Chae, *Considerations For Attorneys Using Artificial Intelligence*, LAW360 (Feb. 14, 2018), https://www.law360.com/articles/1009857/considerations-for-attorneys-using-artificial-intelligence.

[34] Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 88-89 (2014) ("[E]ven given current limitations in AI technology as compared to human cognition, such computational approaches to automation may produce results that are 'good enough' in certain legal contexts.")

[35] *See* HBR CONSULTING, *Artificial Intelligence: Innovations in the Legal Profession*, at 3 (2016), *available at* http://www.hbrconsulting.com/wp-content/uploads/2017/02/Artificial-Intelligence-Resource-Guide.pdf.

[36] *See generally* LEX MACHINA, https://lexmachina.com/ (last visited Feb. 13, 2018). Patent law is an insulated, highly technical area of law, with all binding precedent coming from the Supreme Court and the Federal Circuit, which has appellate jurisdiction over all district-court patent decisions. These unique features of patent law lend themselves to being modeled via predictive analytics, features that are not present in other areas of law.

[37] *See generally* LEXPREDICT, https://www.lexpredict.com/ (last visited Feb. 13, 2018).

[38] *See generally* KIRA, https://www.kirasystems.com/ (last visited Feb. 13, 2018).

[39] *See generally* THOUGHTRIVER, https://thoughtriver.com/ (last visited Feb. 13, 2018).

[40] Ian Lopez, *Will AI Rush in a "Skills Renaissance" in Law?*, LEGALTECH NEWS (Oct. 20, 2016), https://www.law.com/legaltechnews/almID/1202770372162/Will-AI-Rush-in-a-Skills-Renaissance-in-Law/?Mcode=0&curindex=0&curpage=ALL. For example. in Northern Ireland, Ulster University has founded the Legal Innovation Centre, where computing and law students work together to promote advances in access to legal services.

information is most relevant in a given legal matter. With e-discovery, software automatically searches ESI to detect and organize documents.[41] Many large firms have begun to set up internal e-discovery units, partly due to pressure from small, independent firms such as Modus[42] and CloudNine[43] which market themselves as less-expensive e-discovery alternatives. Indeed, the rise in popularity of legal process outsourcing (LPO) has resulted in the "unbundling" of certain legal services such as discovery, a trend that parallels the unbundling of cable TV packages due to the convenience and lower costs offered by online streaming options.[44] In response, some law firms are adopting LPO practices—including a greater reliance on metrics and legal project management (LPM)—while others have built relationships with existing LPO providers.[45]

Going a step further, technology-assisted review (TAR) uses predictive analytics and machine learning to develop more precise and interactive searches. Standard keyword searches run the risk of being both over- and underinclusive: keywords may be absent from relevant documents yet present in other, irrelevant documents.[46] To remedy this issue, TAR identifies not just keywords but "semantics"—word clusters and key concepts—and categorizes the results based on priority (i.e., the likelihood a document, email, etc. contains information relevant to what the user is searching for).[47] Several prominent TAR-based search programs include Judicata (allows lawyers to screen cases for specific procedural and factual details that make them more or less compelling as precedents),[48] NexLP Story Engine (autonomously categorizes structured and unstructured data via concept searches, enhanced threading, and in-depth filtering),[49] Recommind OpenText Axcelerate (uses built-in visual analytics and relevance rankings to identify the "most important facts" for litigation, compliance, and governance purposes),[50] and Catalyst Insight Predict (relies on continuous active learning (CAL) protocols to find and prioritize documents more quickly, and at a lower cost, than traditional TAR engines).[51]

Through TAR, firms are reducing discovery costs while simultaneously discovering the most persuasive, highest quality documents to focus on as they prepare their legal arguments and litigation strategy. While TAR is still an imperfect tool, it is a far more precise and effective means of reviewing documents than relying on simple keyword searches or human efforts alone (with computers not susceptible to distraction, boredom, or exhaustion).[52] The accuracy of machine intelligence will reduce, in some cases, the value of a lawyer's assessments, perhaps forcing the

[41] IAALS, *Navigating the Hazards of E-Discovery, 2d ed.*, at 2 (2012).
[42] *See generally* MODUS, http://discovermodus.com/ (last visited Feb. 13, 2018).
[43] *See generally* CLOUDNINE, https://www.ediscovery.co/ (last visited Feb. 13, 2018)
[44] Ron Friedmann, *The Impact of Legal Process Outsourcing (LPO) You Might Not Have Noticed*, LAW PRACTICE TODAY (Jan. 2012), *available at* https://www.americanbar.org/content/dam/aba/publications/law_practice_today/ the-impact-of-legal-process-outsourcing-you-might-not-have-noticed.authcheckdam.pdf.
[45] *See id.*
[46] John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services*, 82 FORDHAM L. REV. 3041, 3047 (2014).
[47] *See, e.g.*, David van Dijk et al., *Who is Involved? Semantic Search for E-Discovery*, UNIVERSITY OF AMSTERDAM, at 2 (2013), available at https://www.umiacs.umd.edu/~oard/desi6/papers/vanDijk-final.pdf.
[48] *See generally* JUDICATA, https://www.judicata.com/ (last visited Feb. 13, 2018).
[49] *See generally* NEXLP, https://www.nexlp.com/story-engine-description/ (last visited Feb. 13, 2018).
[50] *See generally* RECOMMIND OPENTEXT, https://recommind.opentext.com/products/axcelerate-ediscovery/ (last visited Feb. 13, 2018).
[51] *See generally* CATALYST, https://catalystsecure.com/products/catalyst-insight (last visited Feb. 13, 2018).
[52] John O. McGinnis & Russell G. Pearce, *supra* note 46, at 3049.

lawyer to rely on her "human touch" to supplement the computer's predictions.[53] Ultimately, the extent to which AI and machine learning will replace flesh-and-blood lawyers is an open question. Technology is opening up the law to a vast segment of the population, giving non-lawyers the tools to take care of their legal needs themselves. To stay relevant, lawyers may have to focus on services and skills that computers cannot effectively emulate (e.g., services that require highly specialized knowledge or depend on person-to-person interactions).[54]

IV. Dangers of Using Big Data and Artificial Intelligence

While machine learning is becoming increasingly sophisticated, it is still susceptible to human error. Humans oversee the process in which information is entered into computer algorithms, oversight that can be based on mistaken assumptions. An overreliance on AI could lead employers to naïvely treat these computer projections as "sure things," resulting in the implementation of costly and unnecessary changes to their business plans. Such assumptions are especially problematic in situations where the available data is piecemeal, incorrect, or otherwise misleading. The potential dangers and drawbacks of using big data highlight the importance of knowing *where* the data is coming from and *how* it is being processed.

For example, in situations where the information source itself is flawed, each subsequent step in the data-analysis process is inevitably tainted. Since the reliability of this process hinges on both the quality and quantity of the underlying data, employers should avoid drawing conclusions based only on the numbers themselves, keeping in mind that a particular result may be aberrant or have multiple explanations. There is also concern that humans are asking questions that may not translate well when fed into algorithms: the questions may not be specific enough, or they may presuppose a strong relationship among certain data sets without considering whether the relationship is meaningful or even exists in the first place.[55] Failure to understand the deeper context of a relationship makes it impossible to predict which external factors will affect the relationship, how long the effect will last, and how long the algorithm's results will remain valid.

A. Does Bad Data Create Liability?

When raw data contains errors, or when human error results in faulty data collection or analysis, an employer's reliance on the analysis and underlying data could lead it to make unsound decisions, opening the door to serious legal liability. Little precedent on this issue exists so far. However, the employment sphere seems particularly vulnerable: its heavy reliance on big data could clash with federal and state laws that delineate what information may and may not be considered in the context of employment decision-making. These liability concerns are amplified further when cybersecurity is thrown into the mix, as employers must account not only for the information they are collecting but also for the steps they are taking to protect it, especially when the information in question is confidential or highly personal.

A common misconception among employers is that the non-inclusion of demographic information in a "facially neutral" algorithm will allow all job applicants to receive fair and equal

[53] *See id.* at 3065.
[54] *Id.* at 3065-66.
[55] Darrell S. Gay & Abigail M. Lowin, *supra* note 6, at 9.

consideration for an open position.[56] In reality, an employer's reliance on such an algorithm during the selection process is a recipe for disparate impact liability, where minority applicants are disproportionally overlooked and excluded due to the inadvertent creation of filter bubbles in search engines and information portals. If an employer screens applicants based on her current workforce, and this workforce lacks sufficient diversity, any algorithm used only will magnify the appearance of homogeneity and, in turn, the implication of discrimination.[57] Furthermore, while an algorithm can learn both the characteristics of various protected groups and which traits indicate promising job applicants, it may struggle to determine how much weight to assign these factors. Since algorithms are usually developed by third parties, issues might also arise with respect to indemnification and contributory liability.[58] To defend against allegations of bias, the employer must be able to articulate why one applicant was hired instead of another, which means carefully documenting every step of the hiring process and the reasoning behind key decisions.

B. (Lack of) Formal Regulation

While several U.S. government agencies have noted the risks associated with big data (mis)use, statutory and regulatory sources offer scant formal guidance. Governments around the world are trying to strike a balance between promoting innovation and establishing safeguards to prevent excessive or undisclosed intrusions. As a result, the current regulatory regime has taken a segmented approach, focusing on either the industry as a whole or the specific data collector or handler. There are three general themes: (i) regulate the point(s) of collection ("where is the data gathered from?"), (ii) treat data differently based on public versus private use ("how is the data used?"), and (iii) protect individual privacy through the anonymization or aggregation of personal identifiable information (PII) ("how is the data protected?").[59] Consequently, employers are stuck having to rely on assorted domestic and international regulations (many of which offer scant precedent), nonbinding industry guidelines such as AdChoices,[60] and their own best judgment.

However, employers must still diligently monitor case law and legislative developments that implicate the use of certain data. To encourage ex-convict hiring and reintegration into the workplace, at least nine states plus the District of Columbia have adopted ban the box laws delaying when private employers can ask about a job applicant's criminal history.[61] In addition, to address the gender pay gap, some states and jurisdictions—most notably, Massachusetts—have

---

[56] Braden Campbell, *The Hidden Dangers of Big Data in Employment*, LAW360 (Nov. 17, 2016), *available at* https://www.law360.com/articles/863543/the-hidden-dangers-of-big-data-in-employment.

[57] The Obama administration published several reports warning of big data's impact on employment discrimination, observing how algorithms have the potential to "encod[e] discrimination in automated decisions." *See, e.g.*, The White House, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

[58] *See* Darrell S. Gay & Abigail M. Lowin, *supra* note 6, at 10.

[59] *See* Marcus Evans & Mike Rebeiro, *supra* note 3, at 5. Anonymization plays a critical role in data protection compliance. Once data has been made anonymous, many international data protection rules cease to apply, such as the Data Protection Act 1998 (DPA) in the U.K.

[60] *See generally* ADCHOICES, http://youradchoices.com/ (last visited Feb. 13, 2018). The program calls for online ad companies to establish and enforce responsible privacy practices for interest-based advertising.

[61] Michelle Matividad Rodriguez & Beth Avery, *Ban the Box: U.S. Cities, Counties, and States Adopt Fair-Chance Policies to Advance Employment Opportunities for People with Past Convictions*, NATIONAL EMPLOYMENT LAW PROJECT (May 2017), http://www.nelp.org/content/uploads/Ban-the-Box-Fair-Chance-State-and-Local-Guide.pdf. Nearly thirty states have passed ban-the-box laws that apply to public employers.

passed laws that bar employers from inquiring about an applicant's salary history *at any time* during the hiring process.[62] The growing popularity of these salary history laws coincides with the sharp rise of class actions brought under the Fair Credit Reporting Act (FCRA), which governs an employer's use of an applicant's credit background checks on file with consumer reporting agencies. Recent verdicts in FCRA litigation indicate that both plaintiffs and lawmakers are seeking to crack down on inaccurate credit reports, especially in instances where such reports effectively "defame" the plaintiffs by ruining their employment prospects.[63]

V. Cybersecurity Concerns

Part of the appeal of working with big data is its anonymity. However, the ability to steal and "re-identify" data via PII (birthdates, email addresses, social security numbers) or protected health information (PHI) (medical records) raises serious privacy concerns, especially in the context of cybersecurity.[64] In 2017, data breaches cost upwards of $7 million, with the average size of each breach rising to 24,000 records (in other words, $141 per lost/stolen record).[65] High-profile cyber-attacks over the past few years have targeted not only private companies (Equifax, Uber, Target, and Sony Pictures) but even governments and politicians—the Panama Papers leak and the Democratic National Committee (DNC) hack, to name a few. In the wake of these attacks, which in some cases have resulted in civil liability and other fallout for the hacked entities,[66] establishing comprehensive safety measures to protect collections of big data has never been more vital. Business models that depend on creating and exploiting big data need to develop approaches to information governance that can address—and quickly adapt to—the risks presented by data collection, retention, and analysis.

A. Professional Duty to Protect Client Data

The legal profession is just as vulnerable to cyber-attacks as the business community, if not more so. Evidence suggests that law firm breaches are underreported, leading many lawyers to underestimate the threat they pose and thus not adopt sufficient data-protection precautions.[67] However, the ABA Rules of Professional Conduct—adopted in many states, in whole or in part—specify that lawyers are ethically required to (i) protect confidential client information[68] and (ii) stay up-to-date on "benefits and risks associated with relevant technology."[69]

---

[62] *See, e.g.*, MASS. GEN. LAWS ch. 149, § 105A(c)(2) (2016) (limiting employers to the use of market surveys and prohibiting retaliation for an applicant's refusal to disclose her salary history).

[63] *See, e.g.*, *Ramirez v. TransUnion LLC*, No. 3-12-cv-00632 (N.D. Cal. June 20, 2017) (defendant hit with $60 million verdict for conflating plaintiff class members with similarly named terrorists from a government watch list).

[64] *See* Boris Lubarsky, *Re-Identification of "Anonymized" Data*, 1 GEO. L. TECH. REV. 202 (2017). *See also* Jeff Baker, Op-Ed., *Algorithms Meet Adverbs*, WALL ST. J., Mar. 15, 2017, at A17 ("Every comma is one more tiny piece of Big Data that establishes an online identity as distinct as a fingerprint.").

[65] *See* IBM, *2017 Ponemon Cost of Data Breach Study*, https://www.ibm.com/security/data-breach.

[66] *See, e.g.*, Reuters, *Target Settles 2013 Hacked Customer Data Breach For $18.5 Million*, NBC NEWS (May 24, 2017), https://www.nbcnews.com/business/business-news/target-settles-2013-hacked-customer-data-breach-18-5-million-n764031. Hackers stole data from 40 million credit and debit cards belonging to shoppers who visited Target during the 2013 holiday season. The alleged total cost of the breach was over $200 million.

[67] CNA PROFESSIONAL COUNSEL, *Safe and Secure: Cyber Security Practices for Law Firms*, at 2 (2015), *available at* https://www.aipla.org/resources2/Documents/Safe_Secure_Cyber_Security_Practices.pdf.

[68] ABA Rule 1.6.

[69] ABA Rule 1.1, Comment 8 (part of the duty of competence).

In turn, data thieves both domestically and abroad are eying law firms as easy marks, from which they can readily steal details on prospective corporate mergers, intellectual property (patent and trade secret), litigation strategy, client PII and PHI, and other high-value information.[70] Firms are experiencing thousands of cyber-attacks every day, and several have already "suffered some sort of data breach."[71] These attacks will only intensify as the use of big data becomes more prevalent across the profession. Recognizing this risk, companies are subjecting the cybersecurity policies of their outside counsel to heightened security audits, lest hackers use the outside counsel as a back door to access sensitive corporate information.[72]

B. Cyber Threats and Potential Responses

In response to these attacks, lawyers should familiarize themselves with strategies to defend their collections of big data. Email encryption is one of the most straightforward and effective ways to protect privileged information contained in attorney-client communications, yet many lawyers rely solely on boilerplate "Confidentiality Notices" placed in the subject line or at the bottom of their emails.[73] Notably, one of the top methods used by cyber-criminals to steal data involves email as well: spear-phishing emails contain attachments that, when opened, infect an entire computer network with malware.[74] Other cyber threats include ransomware (holds a victim's files hostage, forcing her to buy a special key to regain access) and "hacktivist" groups (harass firms involved in controversial cases).[75] To get up-to-date information about potential cybersecurity threats, firms should consider joining the Legal Service Information Sharing and Analysis Organization (LS-ISAO), which provides a platform for member firms to share information with each other anonymously.[76]

Even the strongest data-security protections are vulnerable to skilled hackers. A firm should work with its CDS in taking necessary steps to combat the hackers' efforts:

- Create and disseminate a big data policy that will operate alongside the firm's privacy policy and regulate the internal use of collected or imported data sets (e.g., how to use data from public sources, and whether to obtain fair processing warranties for personal data sourced from third parties).
- Have a data breach risk management plan in place, allowing for a rapid response to breaches upon their detection. Many states impose affirmative obligations with respect to protecting client PII and reporting data breaches involving PII.[77] The

[70] Karen Painter Randall & Steven A. Kroll, *Getting Serious about Law Firm Cybersecurity*, NEW JERSEY LAWYER, at 54 (June 2016).
[71] *Id.*; *see, e.g.*, Camilla Hodgson, *Law firm cyber breaches could result in huge thefts and insider trading*, Business Insider (Oct. 19, 2017), http://www.businessinsider.com/cyber-crime-law-firms-vulnerable-2017-10 ("In 2016, insurance company QBE estimated that hackers had stolen £85 million from British law firms over 18 months, after learning they tend to make bank transfers on Fridays and posing as lawyers or clients.").
[72] CNA PROFESSIONAL COUNSEL, *supra* note 67, at 4.
[73] *Id.* at 5.
[74] Karen Painter Randall & Steven A. Kroll, *supra* note 70, at 55.
[75] *Id.* at 55-56.
[76] *See generally* LS-ISAO, https://www.grfederation.org/Members-LS-ISAO (last visited Feb. 13, 2018).
[77] CNA PROFESSIONAL COUNSEL, *supra* note 67, at 3.

breach response team should be familiar with these requirements, particularly if the breach affects a "national" (i.e., multistate) client.[78]

- In addition, have a business recovery/continuity plan in place to help the firm quickly rebound from breaches and other data-related disasters (e.g., having back-ups of important data and keeping these back-ups at a secure, off-site location).
- Regularly conduct vulnerability and malware scans of third-party software programs used within the firm. Relatedly, the firm should vet all third-party vendors of non-legal services (e-discovery, IT, etc.) and ensure its vendor contracts include provisions addressing data treatment and security.[79]
- To guard against the actions of rogue or careless employees, consider adopting data lost protection (DLP) systems, which monitor and limit access to critical data.[80]
- Purchase stand-alone cyber liability insurance, since a number of data breach-incurred costs will not be underwritten by a firm's standard coverage.[81]
- Most importantly, train the firm's lawyers, administrators, and other employees vigorously in how they should protect client data. Besides email and device encryption, lawyers should be trained to change passwords often, maintain strong wireless protocols, exercise caution when working over unsecured networks, and install software on mobile devices allowing them to be wiped remotely.[82]

C. Government Oversight

Failure to institute and maintain cybersecurity protections could result not only in client malpractice claims (along with reputational harm and loss of business) but also in lawsuits brought by federal and state regulatory agencies. For example, PHI theft implicates the Health Information Technology for Economic and Clinical Health Act (HITECH), passed in 2009 to update parts of the Health Insurance Portability and Accountability Act (HIPAA).[83] Since a law firm qualifies as a HITECH "business associate," it would be responsible for securing client PHI, notifying affected parties about the breach, and otherwise complying with HIPAA standards.[84] The Office of Civil Rights (OCR), part of the Department of Health and Human Services, enforces compliance with PHI data regulations, and at times has aggressively targeted businesses that failed to secure this information.[85] Another agency tasked with enforcing cybersecurity compliance is the Federal Trade Commission (FTC), which has the authority under the FCRA to prosecute businesses for allegedly inadequate or lax data security standards.[86]

---

[78] Karen Painter Randall & Steven A. Kroll, *supra* note 70, at 56. Currently, 47 states have notification laws.

[79] CT, *Law Firm Cybersecurity*, WOLTERS KLUWER, at 3 (2016), *available at* https://ct.wolterskluwer.com/sites/default/files/Law_Firm_Cybersecurity.pdf. *See also* ABA Rules 5.1 and 5.3 (duty to ensure vendors' conduct is consistent with other professional obligations).

[80] CNA PROFESSIONAL COUNSEL, *supra* note 67, at 3.

[81] CT, *supra* note 122, at 2.

[82] *See* CNA PROFESSIONAL COUNSEL, *supra* note 67, at 5 ("While advantageous for many reasons, Bring Your Own Device (BYOD) policies are risky if appropriate security measures are not taken.").

[83] Karen Painter Randall & Steven A. Kroll, *supra* note 70, at 56-57.

[84] *See* 42 U.S.C. § 17931(a).

[85] Karen Painter Randall & Steven A. Kroll, *supra* note 70, at 57.

[86] *Id.*

In terms of compliance with international law, many countries have data privacy laws that require the data controller to implement appropriate technical and organizational measures to protect PPI and PMI. For example, EU Regulation 2016/679, known as the General Data Protection Regulation (GDPR), introduces requirements of data protection "by design" and "by default" (i.e., having privacy controls inherently built into systems that collect and process personal data, which include sending promotional emails, storing IP addresses, and even uploading an individual's photo to a website).[87] These controls apply to the entire lifecycle of the data, in part to encourage periodic and routine deletion of old data. The EU also regulates the reuse of personal data that originally was collected for a different purpose.[88] Perhaps the best method to avoid liability relating to data use and protection are through effective consent notifications, which provide individuals with a "real choice" of whether to withdraw their consent.[89]

VI. Conclusion

Employers are relying on big data and computer-based analytics to optimize their workforces, provide equal employment opportunities to employees and job candidates, and streamline general business practices. Lawyers are embracing these new technologies as well, which have the potential to dramatically redefine the legal profession. In applying these technologies, however, caution must be exercised to ensure compliance with both current and nascent legal precedent. Employers should remember to use big data *responsibly*, since they may be *held responsible* if the data is collected, utilized, or released improperly.

[87] Marcus Evans & Mike Rebeiro, *supra* note 3, at 4.

[88] *See* Article 29 Working Party, *available at* http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/fi les/2013/wp203_en.pdf). A four-factor test examines (i) the relationship between the original and present purposes, (ii) the reasonable expecatations of the data subjects, (iii) the impact that further processing will have on the data subjects, and (iv) existing safeguards to prevent undue impact on the data subjects, such as anonymizing the base data or aggregating the results.

[89] Marcus Evans & Mike Rebeiro, *supra* note 3, at 5-6.