



# Purchasing AI for eDiscovery:

NEW, NOW, AND NEXT

# What's inside

Introduction .....	3
Advances in AI-enabled analytics tools .....	5
Purchasing AI for ediscovery .....	6
Four game-changing capabilities .....	8
How to measure ROI .....	14
Questions before you buy .....	17
Common AI pitfalls .....	21
Conclusion .....	25
Glossary.....	26
References .....	27

# Document review as we knew it is quickly coming to an end.

Document review as we knew it is quickly coming to an end. Paying attorneys by the hour to review thousands (and sometimes millions) of documents one-by-one is one of the primary reasons that document review is the most expensive phase of ediscovery. And while that linear review workflow was expensive, but still feasible even just five years ago – the ediscovery landscape has changed dramatically over the last several years.

This change is due to a massive global shift in how people communicate, work, and socialize. From landlines to smart phones, mail-in forms to electronic signatures, memos to chat messages – no matter the industry, every company has undergone a digital transformation over the last decade.

**And what is the universal product of all this change? Data. And lots of it.**

All those new applications and platforms that make our day-to-day easier also billow out a constant stream of more (and increasingly diverse and dynamic) data. The downstream ediscovery effect on law firms and organizations is larger and more complex datasets today than ever before. And those increasingly complicated, larger data volumes are not a trend so much as a sea change. Organizations will continue to create more and more data that will in turn lead to increasingly more diverse, dynamic, and complex ediscovery datasets.

Analytic tools were introduced to infuse efficiency into the ediscovery document review process. However, many of the most common ediscovery analytic tools on the market today still require substantial amounts of linear review by attorneys. Unfortunately, this model is simply too expensive, inefficient, and unreliable to keep up with the pace, volume, and complexity of modern datasets.

**Thankfully, just as technological advances have created this problem, technological advances are also key to managing it.**

# Advances in AI-enabled analytics tools

Improvements in artificial intelligence (AI) technology have given rise to more advanced ediscovery tools that can help firms and organizations tackle mountains of complex data more accurately, efficiently, and cohesively than ever before. For example, tools that leverage this advanced AI technology can more accurately identify and classify targeted data (like responsiveness and privileged information), parse multiple formats, handle today's data volumes, and even provide attorneys with data insights gleaned from an organization's entire legal portfolio. This saves tremendous time and effort for attorneys, lowers overall ediscovery costs, and provides a way to make more strategic decisions that are backed by data – allowing attorneys to get back to what they are trained (and paid) to do: providing legal counsel based on all the relevant evidence.



**Saves time and effort for attorneys**



**Lowers overall discovery costs**



**Make data-based strategic decisions**

# Purchasing AI for ediscovery

So how can your firm or organization take advantage of these developments in AI technology and keep pace with today's modern ediscovery datasets? How do you choose the right advanced AI tools, and what features are right for your organization? And how do you make the case for purchasing new AI-based ediscovery tools to other decision makers?

These are all tough questions, and if you've researched AI tools before, you know that finding the answers can be a grueling experience. Product pages tend to make a lot of broad claims that don't provide an explanation about what the technology actually does. And technology developers often elaborate on the technical functions and performance in ways that don't make sense to the rest of us. To make matters even more complicated, technology providers may generally refer to the technology simply as "AI," even though there are many different subsets of AI with vastly different capabilities. This muddiness around AI may easily lead to ending up with a tool that doesn't meet your needs, which makes it tempting to ditch the effort altogether.

## Don't be dissuaded though – we're here to help.

This buying guide will help you understand what to look for when evaluating a prospective AI tool for ediscovery and help you answer these questions with confidence. We've also included a set of questions to ask ediscovery technology and service providers, which you can print and use when you're researching tools and talking to sales reps.

# Four game-changing capabilities

AI technology has advanced significantly over the last few years. There are now more mature subsets of AI technology that can significantly enhance the practice of ediscovery, especially around document review.

In this section, we've outlined four of the most useful advanced AI capabilities available in ediscovery tools today. Getting a handle on these capabilities will help you choose the best AI solution to tackle the specific challenges your firm or organization faces. Although we've outlined these capabilities separately, leveraging all four of these capabilities together results in a holistic tool that can drive incredible efficiency and accuracy, and up-level entire ediscovery programs.



# Natural Language Processing capabilities

## WHAT IT IS

Natural Language Processing (NLP) enables a tool to understand text the same way that humans understand it. It combines rule-based modeling of the human language with statistical machine learning and deep learning models to process language and “understand” its full meaning. This includes the ability to interpret the intent and sentiment of the writer.

## WHAT IT MEANS IN PRACTICE

NLP enables analytic tools to understand language in context. When leveraged together with deep learning capabilities (outlined on [page 11](#)), newer AI tools can classify language much more accurately than legacy tools. For example, the word “train” could be used in phrases like, “I am going to the train station,” and “I am going to train at the gym.” A tool with NLP used in conjunction with deep learning capabilities can understand the context and then classify those phrases differently, whereas search terms or older AI tools will classify the terms the same.



## PRIVILEGE IDENTIFICATION

NLP's interpretive power pays off in many scenarios, but one of the most helpful is in reviews for attorney-client privilege. This is one of the most expensive phases of document review, because it requires experienced, highly paid attorneys to verify whether documents that are flagged as privilege by an AI tool, privilege screen, or review attorney are indeed privileged. Without NLP, this process can take weeks and cause costs to balloon, because older tools and processes routinely misinterpret text and pull in many false positives. This may lead to high-level attorneys wasting hundreds of hours of expensive billable time reviewing documents that NLP could have correctly identified as not privileged in the first place. Conversely, legacy privilege identification methods can also fail to pull in documents that are privileged, leading to inadvertent privilege disclosures that can increase costly motion practice and divulge sensitive information to opposing counsel. AI-based tools with NLP technology are much better at recognizing privileged language in each individual case – leading to less high-level attorney review, more efficient workflows, and less risk of producing privilege information.

## RESPONSIVENESS IDENTIFICATION

In a similar way, the enhanced classification accuracy provided by NLP can also help greatly reduce the volume of documents pulled into responsiveness pools, which reduces the downstream tasks and checks that are required to be performed for documents that are designated as responsive. Tools without NLP, on the other hand, will sweep in large volumes of non-responsive documents into responsive sets, creating additional layers of time, analysis, and cost related to those downstream tasks. This means that the time and expense of completing those tasks (i.e., tasks like the additional review and redaction of documents containing privilege, highly confidential, and personally identifiable information (PII), as well as the translation of foreign language documents) will have been wasted on documents that shouldn't have been produced in the first place.

# Deep learning processing capabilities



## WHAT IT IS

Deep learning is an advanced form of machine learning that gives AI the power to automate prediction processes within modern datasets. Machine learning uses algorithms to process data and create outputs, but its learning is limited, and it requires people to periodically reprogram it to keep up with changes. Deep learning requires less maintenance because it has multiple layers of algorithms working together to make decisions, and it continues to get smarter with larger and larger volumes of data.<sup>2</sup>

## WHAT IT MEANS IN PRACTICE

Tools with deep learning capabilities are built to handle “big data” – the massive, complicated, and diverse datasets that are common in ediscovery today. Older analytic tools using basic machine learning aren’t up to that challenge. Their “thinking” power is based on a training set, which the tool keeps following until it’s trained again. For example, if data is added or removed in a technology assisted review (TAR) workflow, the TAR model must be reset to align with the new dataset. This can be devastating to matters with tight production deadlines and evolving datasets, such as Hart-Scott-Rodino (HSR) Second Requests.

With AI tools that leverage deep learning, the TAR process can start earlier and continue working as the dataset changes. In short, AI tools capable of deep learning can scale along with any data volume, finding smarter ways to categorize and filter without anyone stepping in to provide additional training or help. An AI-assisted process not only saves time and money, it also enables attorneys to leverage TAR workflows in matters where it would not have been previously possible.

Moreover, when deep learning is combined with NLP, analytic tools can create richer feature sets for the predictive model to learn from, ultimately delivering more accurate results – even from a static dataset.



# Data reuse capabilities



## WHAT IT IS

Data reuse is the process of recalling and learning from old data sets. In traditional ediscovery, legal teams start from scratch on each new matter – regardless of how similar it is to the last one. The same data is reviewed and coded again and again, as if for the first time. This process is clearly inefficient, and as data volumes grow it becomes downright infeasible.

Over the last decade, data reuse has been introduced as a way to minimize this endless cycle of starting anew for each new matter. However, until very recently, this process was limited – consisting of directly porting over attorney decisions and analytic models from one matter to the next in a limited case-by-case basis (generally, in matters with similar custodians and similar responsiveness protocols).

Advancements in AI technology have tremendously enhanced data reuse capabilities. Newer AI tools can ingest, reuse, and learn from previous attorney work, data properties, and multiple models across many past matters – leading to much more accurate analysis than legacy analytic tools.

## WHAT IT MEANS IN PRACTICE

Legal teams can unleash advanced AI to learn from the hundreds of thousands of previously coded documents archived in old databases. Even if documents are not identical from matter to matter, that data and those decisions are still a goldmine of knowledge. This is especially true for classifications that remain relatively the same from matter to matter, like junk documents and sensitive information (e.g., privilege, personally identifiable information (PII), and trade secrets). Advanced AI-enabled analytic tools can learn a lot from analyzing the previous attorney review decisions, as well as past metadata, language use, and other aspects and artifacts. This leads to even more accurate predictions on a wider variety of classifications than ever possible before.



## Portfolio-level analytic capabilities

### WHAT IT IS

Like data reuse, this capability involves leveraging and learning from previous work to improve a legal department's or law firm's overall efficiency and legal strategy, while lowering costs and minimizing risk. But in this case, it also involves the ability to see all the work product and data trends across an organization's entire legal portfolio.

### WHAT IT MEANS IN PRACTICE

This type of portfolio-level dashboard view of ediscovery work product and data enables legal departments to make much more informed, strategic decisions – both on a specific matter and for the organization as a whole. It can help identify trends and make decisions that improve efficiency and strategy, and reduce legal costs and risks. For example, a legal team might find that particular custodians consistently have large volumes of attorney-client privilege or PII within their data. In the short-term, this can help case teams plan for matter costs and make more informed ediscovery burden arguments to courts and opposing counsel. In the long-term, it can better inform information governance and data retention policies. A full portfolio view can also assist case teams even before collecting data for a current matter, by giving insight into what type of information may be residing in a custodial or data source collection and how it was previously coded. This early insight can help immediately inform case strategy and control legal costs. Or an in-house legal team may find patterns regarding increased litigation when certain factors (such as specific custodians or data sources) are involved in a matter. This can help organizations minimize risk and improve workplace compliance in the long term.

# How to measure ROI

As a partner to human reviewers, advanced AI tools provide a powerful return on investment (ROI) in many aspects of the ediscovery process. Understanding this ROI can help you build a convincing business case prior to purchase, as well as help evangelize usage among attorneys and measure success over time.

The following are four primary benefits of using AI-enabled analytics for ediscovery.



## Early insights into matters and faster workflows

The right advanced AI analytic tool will give you a clearer view of your matters at an earlier stage in the game, allowing you to make more informed legal and strategy decisions right from the outset. It will also enable you to complete ediscovery workflows, from responsiveness identification to privilege review, more efficiently. Enhancements in AI technology over the last few years have resulted in advanced AI tools that work faster, even when dealing with larger datasets.

**Imagine the risk mitigation, quality control, and better work product that can occur when every production isn't a fire drill – completed hours or minutes before the deadline due to manual, slower workflows using legacy ediscovery tools.**



## Lower overall ediscovery and review spend

The right advanced AI tool will reduce overall legal costs by improving efficiency and case strategy. Document review is one of the most expensive stages of ediscovery, and an advanced AI analytic tool can greatly reduce the need for eyes-on review.

**In fact, advanced tools that leverage all four of the capabilities outlined earlier can be expected to reduce ediscovery costs by 30% or more.<sup>3</sup> Additionally, seeing trends across matters enables strategic decisions that lead to program-level cost savings.**



## Improved accuracy

Studies have shown that machine learning tools from a decade ago are at least as reliable as human reviewers<sup>4</sup> – and today's more advanced AI tools are even more accurate than they were before. Analytic tools built with AI do a better job of detecting privilege, PII, confidential information, and junk data. This saves a wealth of time, trouble, and risk down the line.

**More accurate responsiveness classifications lead to fewer downstream tasks like redactions, privilege review, and foreign language translation, which ratchet up costs, complications, and timelines. They also lower the risk of inadvertently disclosing non-relevant but sensitive information that could fuel more litigation.**



## Risk mitigation

With better accuracy comes lower risk. But advanced AI tools take this benefit even further by applying their findings across matters: once sensitive information is identified and withheld in one matter, it can be consistently identified that way in future matters – significantly lowering the risk of disclosing sensitive and privileged information to opposing counsel and investigators. Additionally, a dashboard view of an organization's entire legal portfolio enables legal teams to identify risk trends they otherwise would not have. For an example, a higher incidence of litigation across certain custodians, or a trend of outdated material stored in certain data sources, etc.

**Overall, the right AI tool can help mitigate risk on each new matter, as well as across the entire organization.**

# Questions before you buy

Once you know what capabilities to look for and how to measure ROI, you're ready to start comparing tools. The questions on the following pages can help you navigate conversations with technology providers and identify the tools that best fit the needs of your organization or firm. Remember that as you are asking providers these questions, you'll also want to ask them to provide concrete examples and quantifiable results for each response. This section is designed as a worksheet for you to reference when talking to providers about AI-based analytic tools.

USE CASES	
General	<p>What are the possible use cases for the tool?</p> <p>Can the provider share real-life examples?</p>
Portfolio-Level Analysis Capabilities	<p>Can the tool be used to gather trends and insights from an entire legal portfolio?</p> <p>How are insights communicated?</p> <p>Is there a visual dashboard, a report on a regular cadence from a service team, or both?</p> <p>Can the provider share examples of where that was done successfully, including info about cost savings and improved efficiency?</p>
Cross-Matter Capability	<p>Can the tool provide analytics and efficiencies across matters?</p> <p>How are matters connected?</p> <p>Can matters be connected if their hash values were processed differently?</p> <p>Can the provider share examples of where cross-matter capability was used successfully?</p>
Review Needs	<p>Can the tool be used for family-level review, four corner review, or both?</p>
Review Needs: Classification Types	<p>What document review categories can the tool classify (attorney-client privilege, PII, junk data, responsive documents, etc.)?</p> <p>Can the tool handle custom classifications, like trade secret or highly confidential data?</p>
Review Needs: Proven Classification Accuracy	<p>Can the provider share examples of the tool's accuracy in identifying and classifying a variety of information?</p>
TAR Workflows	<p>Which TAR workflows can the tool be used for (i.e. TAR 1.0, TAR 2.0, etc.)?</p> <p>Can the tool be used effectively in TAR workflows without the need to retrain if the dataset changes?</p> <p>Is the tool's analysis statistically defensible?</p> <p>Can the tool be leveraged for TAR workflows on datasets where traditional analytic tools would normally fail (for example, datasets with extremely high richness or large volumes of foreign language, etc.)?</p> <p>Have government agencies like the Department of Justice (DOJ) or Federal Trade Commission (FTC) accepted the technology in TAR workflows?</p>
Overall eDiscovery Program Improvement	<p>Will the product help my firm/organization build a more effective ediscovery program?</p>

UNDERLYING TECHNOLOGY	
AI Technology Types	<p>Does it use older technology such as logistic regression and support vector machine (SVM) vs. newer technology such as transformer-based algorithms and deep learning?</p> <p>Is the technology leveraging NLP so that it can understand human language?</p> <p>Is the technology leveraging deep learning so that it can more accurately handle modern datasets?</p>
Diverse Data Capability	Is the tool able to parse and analyze data from a variety of data sources (chat messages, Excel spreadsheets, PDFs, emails, etc.)?
Document Feature Analyzation Capability	What document features can the technology analyze? For example, can the tool only analyze the text of a document, or can it analyze a document from multiple angles (text, metadata, prior coding decisions, etc.)?
Scalability	What is the largest dataset the tool has been used on (including what the tool was used for and results of that use)?
RESULTS VISUALIZATION AND QUANTIFICATION	
Quantified Results	Can the provider quantify the expected results of the tool, with real case studies and examples?
Client References	Can the provider offer references from clients who have successfully used the tool?
Visualization Capabilities	<p>Is there a data dashboard that presents results in a visual way that is easy to understand for the average user?</p> <p>Is the dashboard easy to access and manipulate?</p> <p>Can users download and share result dashboards?</p>
ONGOING USER SUPPORT	
Training	Is training available for the tool? What does it cost?
Expertise	<p>Does the provider have data and technology experts available to run the tool and/or explain results to internal stakeholders, outside counsel, opposing counsel, and/or judges?</p> <p>What are the qualifications of those experts?</p> <p>Does the provider have legal and technology experts that can testify in court or provide expertise in a variety of settings, for example as a (30)(b)(6) witness, in meet and confers, or in negotiation with government agencies or opposing counsel, etc.?</p> <p>What are the qualifications of those experts?</p>
Users	Must users learn to use the technology themselves or can the technology provider leverage the technology at the direction of case teams, and simply supply results?
Technical Support	<p>Does the provider offer technical support for the tool?</p> <p>What is the associated time and cost?</p> <p>Does the supplier have in-house data scientists available to clients?</p>

TIMELINE	
Matters	<p>How much work will it take to implement on a particular matter?</p> <p>How long will it take to first see results?</p> <p>How long will it take to see all results in a matter?</p>
Results	How long will it take to see classification results from when the data is ingested?
Training the Model	How many examples does it take to train the model?
Benefits	How long will it take to see benefits from a portfolio-level analysis?
INTO THE FUTURE	
Road map	What are the provider's future plans for the tool (i.e., the "development roadmap")?
Technology Investment	<p>How does the provider plan to invest in AI tools in the future?</p> <p>Does the provider invest in and build its own technology, or does it acquire it from other companies?</p>
eDiscovery Innovation	<p>How has the provider invested in ediscovery?</p> <p>Is the provider familiar enough with the industry to be able to innovate as needed?</p>
ONBOARDING AND INTEGRATION	
Integration	<p>Is the AI tool integrated into the provider's review platform, or will it need a separate implementation?</p> <p>If the latter, what ediscovery platforms can the tool be integrated into?</p>
Data Storage and Security	<p>Where is data stored?</p> <p>How does the provider ensure that data is secure and meets all applicable data protection requirements?</p>
Data Migration	<p>How is data migrated into the tool?</p> <p>Can data processed by different providers be connected for maximum ROI?</p> <p>Can it accept data from any platform? Most companies have had data sit in multiple platforms from different providers.</p>

# Common AI pitfalls

If not implemented properly, even the best technology will fail to return the desired results. Avoid the following pitfalls to ensure that your firm or organization is maximizing its investment in the AI tool you select.



## Being too rigid about workflow process changes

To achieve the best results with AI-based tools, some review workflow changes may be necessary. For example, many attorneys may be most familiar with a “family-level” review process, where reviewers decide how to code a document based on the content of other documents in the same family. However, a “four corner” review process works best for AI-based tools, which analyze each document individually. For example, an AI-enabled analytic tool might code a blank attachment as junk, even though an attorney coded it as responsive and privileged in family-level review because it was attached to a responsive and privileged email. Additionally, the ability to reuse or propagate coding across matters is only possible with a “four corner” review process.

**When attorneys or case teams are too rigid or unwilling to adapt to review process changes, they will lose out on some of the biggest benefits of advanced AI tools.**



# 2

## Not trusting or fully leveraging the technology

There is no better way to destroy an AI tool's ROI than to use it solely as a backup to human review (i.e., an additional layer of quality control). If you don't trust the tool to make some decisions on its own – such as recognizing that email signatures and blank pages are not worth a human reviewer's time – you won't get much out of it. Attorneys should select an AI-based tool that enables them to minimize human review in a statistically defensible way.

**This means properly vetting the tool's accuracy and efficacy before purchasing, as well as testing it with a random sample once you bring the tool on board. But after that's done, case teams should let the AI do the grunt work and allow their people to focus on the reviews and strategies that are truly worth their time.**



## Avoiding expert consulting

Attorneys should not be expected to be data scientists. A quality technology provider will have data scientists, linguists, and AI experts, whose jobs are to help clients maximize their investment in the technology and ensure case teams have the support they need on each case.

**Experts can step in to fully implement the tool on any given case, explain workflows and results (including providing expert testimony), and provide insight into how the tools can be used to maximize investment across matters and legal portfolios.**

# Conclusion

Today's rapidly evolving data landscape has made document review in ediscovery more challenging, more expensive, and more time consuming than ever before. Advancements in AI technology are not just the key to overcoming those challenges – they are also the key to effectively leveraging ediscovery data to make better, more strategic business and legal decisions.

With this buyer's guide, you'll be better prepared to find a technology partner that can provide the most advanced AI technology that fits your firm or organization's unique needs.

**Don't drown in a sea of data. Embark on the next leg of your ediscovery journey with the right AI tool.**

# Glossary

The new normal for ediscovery includes several new terms. Having a basic understanding of these terms can help you talk to technology providers, research your options, and explain them to others at your firm or organization.

- **Artificial Intelligence (AI):** Broadly, the “science and engineering of making intelligent machines, especially intelligent computer programs.” IBM clarifies that AI is “a field, which combines computer science and robust datasets, to enable problem-solving. ” AI includes subfields of machine learning and deep learning.<sup>7</sup>
- **Big Data/Modern Datasets:** Datasets with such high volume, high velocity, and/or high variety that traditional relational databases are unable to capture, manage, and process their contents. These datasets are more complex than traditional datasets due to a variety of modern-day influences, including AI, mobile devices, social media, and more.<sup>8</sup>
- **Deep Learning:** A subset of machine learning that does not require human intervention to process data. It consists of a neural network with “three or more layers,” which simulate the behavior of the human brain and learn from large amounts of data. Because they have multiple layers, the predictions from deep learning models are more refined than those of machine learning models.<sup>9</sup>
- **Logistic Regression:** Logistic regression is a classification method that models the probability of a discrete outcome. It is a simple and efficient method to solve binary problems – for example, problems with two values outcomes such as true/false, yes/no, etc.<sup>10</sup>
- **Machine Learning:** A branch of AI that “focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.”<sup>11</sup> In machine learning, algorithms are trained to make classifications or predictions.<sup>12</sup> This is especially important in ediscovery as data volumes continue to grow and diversify. Traditional machine learning requires human intervention to process data.<sup>13</sup>
- **Natural Language Processing (NLP):** A branch of AI computer science that is concerned with giving computers the ability to understand text (and spoken word) in the same way that humans understand it. It combines rule-based modeling of the human language with statistical, machine learning, and deep learning models to process human language and “understand” its full meaning, including the intent and sentiment of the writer (or speaker).<sup>14</sup>
- **Supervised Machine Learning:** A subcategory of machine learning that uses labeled datasets to train algorithms to classify data or predict outcomes. As data is inputted into the model, machine learning weights that data until the model “stabilizes.”<sup>15</sup>
- **Support Vector Machines (SVM):** A widely used supervised machine learning classification technique, developed in 1995. It is based on statistical learning theory with the goal to project nonlinear separable samples onto another higher dimensional space by using different types of kernel functions.<sup>16</sup>
- **Transformer Based Algorithms:** A deep learning model that has achieved state-of-the-art results by differentially weighting the significance of each part of the input data.<sup>17</sup>
- **Unsupervised Machine Learning:** A subcategory of machine learning, where algorithms can analyze and cluster unlabeled datasets to discover patterns or data groupings without the need for human intervention.<sup>18</sup>

# References

- 1 "Machine learning" is defined in the glossary here. Deep learning is described in more depth on the [previous page](#).
- 2 IBM Cloud Education. "Deep Learning." May 1, 2020. <https://www.ibm.com/cloud/learn/deep-learning>
- 3 "Lighthouse Prism Reduces Review Costs Across Client's Entire Legal Portfolio." <https://assets2.brandfolder.io/bf-boulder-prod/f9672zk9sn6vx44nswbx7mq/v/49580686/original/Lighthouse%20Prism%20Reduces%20Review%20Costs%20Across%20Clients%20Entire%20Legal%20Portfolio.pdf>
- 4 Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, 17 Rich. J.L. & Tech 11 (2011). Available at: <http://scholarship.richmond.edu/jolt/vol17/iss3/5><https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1344&context=jolt>
- 5 McCarthy, John. "What Is Artificial Intelligence." Standard University Computer Science Department. [https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems\\_2008\\_2009/Old/IntelligentSystems\\_2005\\_2006/Documents/Symbolic/04\\_McCarthy\\_whatissai.pdf](https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_whatissai.pdf)
- 6 IBM Cloud Education. "Artificial Intelligence (AI)." June 3, 2020. <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>
- 7 Id.
- 8 IBM. "What is big data analytics?" <https://www.ibm.com/analytics/hadoop/big-data-analytics>
- 9 IBM Cloud Education. "Deep Learning." May 1, 2020. <https://www.ibm.com/cloud/learn/deep-learning>
- 10 Science Direct. "Logistic Regression"; Chapter 4 - Exploratory Study, Thomas W. Edgar, David O. Manz, in Research Methods for Cyber Security, 2017. Abdulhamit Subasi, in Practical Machine Learning for Data Analysis Using Python, 2020. <https://www.sciencedirect.com/topics/computer-science/logistic-regression>
- 11 IBM Cloud Education. "Machine Learning." July 15, 2020. <https://www.ibm.com/cloud/learn/machine-learning>
- 12 Id.
- 13 IBM Cloud Education. "Artificial Intelligence (AI)." June 3, 2020. <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>
- 14 IBM Cloud Education. "Natural Language Processing (NLP)." July 2, 2020. <https://www.ibm.com/cloud/learn/natural-language-processing>
- 15 IBM Cloud Education. "What is Supervised Learning." August 19, 2020. <https://www.ibm.com/cloud/learn/supervised-learning>
- 16 Science Direct. "Support Vector Machines". Derek A. Pisner, David M. Schnyer, in Machine Learning, 2020. Sandeep Kumar Satapathy, ... Shruti Mishra, in EEG Brain Signal Classification for Epileptic Seizure Disorder Detection, 2019, <https://www.sciencedirect.com/topics/neuroscience/support-vector-machine>
- 17 Towards Data Science. "Transformer in CV". December 7, 2020. <https://towardsdatascience.com/transformer-in-cv-bbdb58bf335e>
- 18 IBM Cloud Education. "What is Unsupervised Learning." September 21, 2020.



## About Lighthouse

For 25 years, Lighthouse has provided innovative software and services to manage the increasingly complex landscape of enterprise data for compliance and legal teams. Lighthouse leads by developing proprietary technology that integrates with industry-leading third-party software, automating workflows, and creating an easy-to-use, end-to-end platform. Lighthouse also delivers unique proprietary applications and advisory services that are highly valuable for large, complex matters, and a new SaaS platform designed for in-house teams. Whether reacting to incidents like litigation or governmental investigations, or designing programs to proactively minimize the potential for future incidents, Lighthouse partners with multinational industry leaders, top global law firms, and the world's leading software provider as a channel partner.

**Contact us to find out what Lighthouse can do for your business.**

206-223-9690 | [lighthouseglobal.com](http://lighthouseglobal.com) | [info@lighthouseglobal.com](mailto:info@lighthouseglobal.com)