WHITE PAPER

TAR + Advanced AI: The Future Is Now

Demystifying TAR and advanced AI to up-level your ediscovery

LIGHTHOUSE

Contents

Introduction

The concept of discovery in litigation dates to Common Law and the English Court of Chancery.[1] However, it wasn't until the late 1990s that the revolutionary rise of computing began to dramatically increase electronically stored information (ESI), catalyzing the first dramatic change to the civil discovery process in decades. The new document types forced lawyers and judges to change age-old discovery and litigation processes to reflect the reality of how people were now storing and exchanging information.

By 2006, attorneys began to grapple with new requirements to preserve and produce ESI during discovery, due to groundbreaking amendments made to the Federal Rules of Civil Procedure (FRCP). Case teams faced with large volumes of ESI were forced to hire big teams of contract attorneys to individually review electronic documents, flagging important information and separately tagging the documents for responsiveness, privilege, confidentiality, etc. Depending on the volume, complexity, and number of attorneys, this process could take months or even years of work, billed out hourly.

Enter more advanced technology. Beginning in the mid-to-late 2000s, case teams started to use a subset of artificial intelligence (AI) technology called "machine learning" to predict the responsiveness of documents, thereby expediting the review process. This process became known as technology assisted review (TAR). As the legal industry's first foray into AI, the term "TAR" was often conflated with the technology behind the tool – even though AI actually encompasses a much broader set of technology. Even today, when more advanced AI tools exist, many attorneys think of TAR technology when they hear the term "AI" used in the ediscovery context. Conversely, other attorneys may think of a myriad of unrelated ediscovery tools that do not, in fact, fall within the definition of AI.

This muddiness around terminology has become a detriment to the important ediscovery work that attorneys perform today. Now more than ever, attorneys must have a solid grasp of how AI technology works to choose the best available tools and meet obligations to their clients and organizations.

That's where we come in. We know not all attorneys are technologists. The goal of this paper is to deepen the reader's understanding of AI within the context of ediscovery and provide answers to common questions. We will focus on TAR to provide a familiar framework with which to talk about applications of advanced AI in ediscovery. In the remainder of this paper, we will outline the history and definition of TAR, define advanced AI, and then delve into the interconnection between TAR and advanced AI, how to choose the appropriate technology, and when to use it.

What is TAR?

To discuss how TAR can be revolutionized by combining it with the power of advanced AI, it is first necessary to level set by defining what TAR is and how it has evolved thus far.

Currently, TAR is performed by machine learning algorithms that classify documents for responsiveness based on human input or "training." This classification allows attorneys to efficiently prioritize the most important documents for eyes-on review. Further, when agreed to by opposing counsel and/or a relevant governing body, the party performing TAR may also be able to avoid reviewing documents that the tool has determined are very unlikely to be responsive. In this way, TAR can significantly reduce the number of documents humans need to review, thus it has historically been helpful (and sometimes critical) when dealing with short deadlines or larger volumes of data.

THE HISTORY OF TAR

TAR as we know it today evolved from a more basic form of data classification performed by machine learning algorithms, which was first outlined in Anne Kershaw's 2005 study entitled "Automated Document Review Proves Its Reliability."[2] In that study, Ms. Kershaw's research found that "electronic relevance assessment application and process reduced the chances of missing relevant documents by more than 90 percent."[3] That research helped pave the way for more widespread acceptance of this type of technology in the ediscovery space.

Between 2006 and 2010, ediscovery technology advanced into the first form of TAR, or what is now referred to as TAR 1.0, with TAR 2.0 following shortly thereafter. This introduction created a buzz in the industry, as it gave litigation teams the ability to handle growing ESI volumes with much more efficiency and a fraction of the cost of manual review.

TAR 1.0

TAR 1.0 uses supervised machine learning, where a small number of highly trained subject matter experts review and code a randomly selected group of documents called a control set. The control set provides an initial overall estimated richness metric and establishes the baseline against which the iterative training rounds are measured. Through the training rounds, the machine develops a classification model.

Once the training rounds no longer improve the classification, the system is considered to have reached stability. At this point, the computer applies scores to all the documents in the dataset, with lower scores indicating documents less likely to meet the criteria set out by the experts in the training session. Using statistical measures, a cutoff point or score is determined and validated, above which the desired measure of relevant documents will be included. The remaining documents below that score are deemed not relevant and therefore do not require any additional review.[4]

As previously noted, it was during the late 2000s that TAR 1.0 began to be used in a limited number of larger document reviews – in part due to influential bodies, such as the Text Retrieval Conference (TREC) and the Sedona Conference, issuing papers and studies on discovery search and text retrieval methods in the electronic discovery space.[5] In 2011, Maura Grossman and Gordan Cormack published a research paper titled, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review."[6] This paper was important in the history of TAR, not only because it evaluated the efficacy and efficiency of the TAR 1.0 technology and methodology as we know it today, but also because it found that TAR 1.0 could actually "yield results superior to those of exhaustive manual review."[7]

In 2012, Judge Peck, then a Magistrate Judge in the Southern District of New York, issued his seminal opinion in *Da Silva Moore v. Publicis Groupe*, approving the utilization of "computer-assisted coding" in federal court where both parties agreed to its use.[8] Importantly, in his decision, Judge Peck quotes an article he wrote about the subject, where he noted that many attorneys were waiting to use TAR until it was approved by a federal court. This decision, he stated, would be that long-awaited approval.[9]

During that same year (2012), the RAND Corporation issued a groundbreaking research brief titled "The Cost of Producing Electronic Documents in Civil Lawsuits," which called out the inefficiency, inaccuracy, and expense of conducting ediscovery the old way (i.e., via manual review). The authors of that brief looked to the "nascent technology" of predictive coding as a possible solution, lauding its consistency and efficiency.[10]

The Da Silva Moore opinion and industry buzz helped open the floodgates for attorneys and their clients to start learning about and using TAR 1.0 in more cases. With more cases, more court decisions followed that approved its use in a variety of different scenarios.

TAR 2.0

As TAR 1.0 took hold, it became clear that it could be useful in specific types of datasets. However, because it used simple machine learning, TAR 1.0 use cases were limited to cases where all documents that needed to be reviewed were available at the outset of the matter. TAR 1.0 was found to be less beneficial in cases where document collection is ongoing or "rolling," because the system must be retrained with each addition. TAR 1.0's requirement for highly trained subject matter experts to train control sets also meant the cost-value ratio was substantial enough to justify its use only in large static data volumes – because those experts' review work was more expensive than regular attorney review.[11]

The second iteration of TAR, TAR 2.0, uses the same supervised machine learning technology as TAR 1.0, but rather than the simple learning of TAR 1.0, TAR 2.0 utilizes active learning. With active learning, the system will continuously learn from reviewer decisions. This means that TAR 2.0 does not require a one-time set of decisions by the subject matter expert team on a control set to train the system at the outset of the matter in the way that the TAR 1.0 process requires. Instead, regular attorney review teams can immediately start reviewing documents in a TAR 2.0 model. This also means that the review set does not need to remain static in TAR 2.0. The software continues to learn as attorneys review and code documents (no matter when those documents are added to the dataset) and will continue to score those documents. In turn, this allows case teams to prioritize their workflow to ensure that the most responsive documents are always being reviewed next.[12]

TAR 1.0 vs. TAR 2.0

Both TAR 1.0 and TAR 2.0 are still widely used in ediscovery today – and both the technology and processes remain relatively unchanged despite more than a decade of utilization. Each has its own advantages and disadvantages depending on the matter parameters.

As previously noted, a TAR 2.0 model can be more beneficial than a TAR 1.0 model in matters where review needs to start immediately, or when attorneys want to introduce TAR to a review that is already underway.[13] TAR 2.0 is also more flexible in implementation and can be used for a variety of different use cases. For example, TAR 2.0's prioritization review workflow (wherein the highest scoring documents are continuously pushed to the front of the review as the technology learns from reviewers, creating a loop where the most up-to-date model identifies which documents should be reviewed next) can be helpful for matters that have short, rolling production deadlines.[14]

TAR 1.0 still has its advantages over TAR 2.0 – especially in its ability to stabilize a model and require fewer overall documents to be reviewed when compared to a TAR 2.0 model. For example, Lighthouse has seen cases where as few as 5,000 documents were required to stabilize a TAR 1.0 model for a 1-million total document population. It can also lead to smaller overall review set in a shorter amount of time.

TAR 1.0 is also still the primary workflow used in regulatory investigations, like Hart-Scott-Rodino Act (HSR) Second Requests (the discovery process by which the federal government investigates potential mergers and acquisitions for anticompetitive behavior) due to several of its advantages over TAR 2.0. The nature of these investigatory discovery matters involves a very short discovery period with large data volumes and rigorous production requirements that, if not met, involve harsh penalties and risks. This means that quickly culling data from a dataset (in a way that is acceptable to regulatory bodies like the Department of Justice (DOJ) or the Federal Trade Commission (FTC)), is critical. TAR 1.0 is particularly effective for this task because it's a much easier workflow to negotiate culling data. With a TAR 1.0 workflow, case teams can negotiate a statistically validated cut-off score with the regulatory body at the outset of the investigation, below which they will not have to produce any documents. Then, once the model stabilizes, the case team can simply produce everything above the negotiated cutoff without further responsiveness review. This contrasts with a TAR 2.0 workflow, which does not involve a control set nor does it involve the standard recall precision statistical metrics that can validate stopping reviewing after a certain point to regulatory body. TAR 1.0's longer history also helps attorneys feel more comfortable with that workflow over TAR 2.0, making its use more prevalent.

However, because the "machine learning" technology used in both TAR 1.0 and TAR 2.0 is older and not built for big data, both processes are quickly becoming limited in their applications with modern datasets.

The future of TAR

At Lighthouse, we believe that TAR is finally on the verge of a sea change. As data volumes continue to explode and become more varied and complex, the supervised machine learning technology behind TAR is becoming inadequate to manage modern data. A decade ago, TAR was generally able to handle that era's data volumes and limited variety of data sources. Email was the standard form of communication, and there was much less volume and diversity within data – meaning that a TAR 1.0 or TAR 2.0 workflow solely utilizing supervised machine learning could be effective enough to accurately classify the most common forms of data at the time, and thereby greatly reduce the amount of time spent on eyes-on review.

But today's datasets are vastly different than they were a decade ago. Today's employees use myriad applications to communicate and work (think: chat systems, smartphones, cloud-based collaboration tools that incorporate a dozen different applications within them, etc.) rather than communicating and working within the limited number of applications that were available 10 years ago (think: databases, email, Microsoft Word and Excel, etc.).

To have true efficacy in this modern environment, systems trained to classify data must be able to handle the additional volume, as well as grasp the different ways information is communicated within constantly evolving data sources. This requires a combination of technologies, heavily leveraging the more advanced subsets of AI available today – specifically, the subsets of deep learning and natural language processing (NLP), both of which are explained in more depth later in this white paper.

Like the original ediscovery and TAR evolutions that came before, if attorneys are not willing to evolve and adapt with technology to manage the realities of modern data volumes and diversification, they risk falling down on their ediscovery requirements and subsequently failing their clients.

Accordingly, we believe that updated TAR processes that utilize advanced AI will soon not only be considered an optional tool for forward-thinking attorneys who want to provide clients with better and more efficient discovery processes, but will also be an ethical duty on the part of attorneys necessary to meet big data ediscovery requirements.

What is advanced AI?

As previously noted, the term "AI" in ediscovery has become muddied – somehow simultaneously bringing to mind the somewhat limited functionality of TAR while conjuring futuristic images of human-like robots and fully self-driving cars. Because AI encompasses a large swath of technology, simply using the term "AI" can result in conflating completely different technologies. Thus, this section will begin by defining advanced AI as it applies to ediscovery and how that technology can be used to improve upon ediscovery – and more specifically TAR 1.0 and TAR 2.0 workflows – in today's more complicated and voluminous datasets.

Broadly speaking, AI refers to the "science and engineering of making intelligent machines, especially intelligent computer programs."[15] It encompasses the subfields of machine learning and deep learning.[16] Machine learning "focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy where those algorithms are trained to make classifications or predictions."[17] This is especially important in ediscovery, as data volumes continue to grow and diversify. Older subsets of machine learning, like supervised machine learning used in TAR, require human intervention to process data.[18] Deep learning, while a subset of machine learning, "automates much of the feature extraction piece of the process, eliminating some of the manual human intervention required and enabling the use of larger datasets."[19] Natural language processing (NLP) is also a separate branch of machine learning that involves training computers 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).[20]

Thus, when we refer to "advanced AI" in this paper, we are referring to a combination of the aforementioned technology that utilizes multiple branches of AI (i.e., traditional machine learning in combination with deep learning and NLP) to make more accurate and efficient predictions within modern datasets (i.e., datasets that are more voluminous and diverse than even five years ago). Below, we've defined a few other technology terms that will be helpful as we delve further into the topic of "advanced AI".

AI TERMS TO KNOW

Artificial Intelligence (AI): Broadly, AI refers to the "science and engineering of making intelligent machines, especially intelligent computer programs."[21] IBM clarifies that AI is "a field, which combines computer science and robust datasets, to enable problem-solving."[22] AI also encompasses subfields of **machine learning** and **deep learning**.[23]

Big Data: Datasets that are beyond the ability of traditional relational databases to capture, manage, and process the data with low latency due to high volume, high velocity, and/or high variety. These datasets are more complex than traditional datasets due to a variety of modern-day influences, including AI, mobile devices, social media, etc.[24]

Deep Learning: Deep learning is 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" that simulate the behavior of the human brain and learn from large amounts of data. Because of its multiple layers of neural networks, its predictions are more refined than machine learning.[25]

Graphical Modeling: A graph that shows the probalisitic relationships among a set of variables. Within the ediscovery space, graphical modeling is often used to show communication networks within a dataset (see social network analysis below).[26]

Limited Memory: Unlike reactive AI, limited memory systems *do* have the ability to learn from historical data to make decisions. Machine learning and deep learning use limited memory to train by ingesting large volumes of data that is then stored in their memory from a reference model for solving future problems.

Machine Learning: Machine learning is a branch of AI, which "focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy."[27] In machine learning, algorithms are trained to make classifications or predictions.[28] This is especially important in ediscovery as data volumes continue to grow and diversify. Traditional machine learning requires human intervention to process data.[29]

Natural Language Processing (NLP): NLP is 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).[30]

Reactive AI: Reactive AI systems are a much older (and somewhat outdated in the non-ediscovery world) system of AI. For example, reactive systems were what IBM's famous "Deep Blue" machine used in the late 1990s. Reactive AI does not have the ability to use previous data and experience to inform present decisions. It can only be used to automatically respond to a limited set of inputs, and thus are not used by machine learning or deep learning.[31]

Social Network Analysis (SNA): Visual representation of social networks and communication, using analytic software and graphical modeling. In the legal space, SNA is extremely valuable for case teams to gain a big-picture view of how people represented within the dataset were communicating (who they were communicating with, how often, what times of day, etc.).[32]

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."[33]

Unsupervised Machine Learning: A subcategory of machine learning where algorithms are able to analyze and cluster unlabeled datasets to discover patterns or data groupings without the need for human intervention.[34]

TAR vs. AI

Common questions around TAR and AI center on their relation: Is TAR, in fact, AI? Does AI in ediscovery mean TAR? The answer is this: The "technology" most commonly behind traditional Technology Assisted Review is supervised machine learning, a subset of AI as previously noted. So yes, the current TAR process uses a subset of AI. However, it's worth noting that the other part of TAR is the process with which the technology is implemented to ultimately classify the responsiveness of documents.

The difference between legacy TAR and TAR with advanced AI lies in the additional capabilities opened up by advanced AI subsets outlined to classify data. The supervised machine learning behind traditional TAR implementation uses raw statistics and text analyzation to drive "feature extraction" (i.e., whatever the machine is trying to predict). However, when an ediscovery tool combines that statistical prediction with deep learning and NLP, it can first learn a language model that is then used to make much more accurate classifications – because the language model enables the tool to understand the meaning of words in the context of others. For example, with traditional TAR, the word "train" in the phrases "I am going to the train station," and "I'm going to train at the gym," are classified the same way – because statistically, there is not much difference between how the word "train" is placed within both sentences. However, newer tools that combine supervised machine learning with deep learning and NLP can learn the context of when the word "train" is used to mean the noun "train," and when the word is being used as a verb.

Additionally, ediscovery tools that leverage deep learning and NLP can analyze more than just the text of the document – they can analyze the document from a variety of angles, including the metadata and data source when processing and learning from data. Thus, rather than throwing all data into a blender and extracting word meanings from statistics, tools that utilize deep learning and NLP can recognize that a word used in a chat platform may produce different results than the same word used over email. This ability can be especially useful in identifying privilege, as attorneys may speak differently over an informal chat system while discussing fantasy baseball with a large group of co-workers than within an email conveying legal advice to one person within the company – even if statistically, the word usage is similar. The context of the data source and how words are used matters, and an advanced AI tool that leverages a combination of technologies can better understand that context.

Given the above, the real conversation can then shift from TAR vs. AI to "what kind of AI is best to use to execute a TAR workflow?" And we'll do just that in the next section where we outline the "when" and "why" of how to apply various subsets of AI in TAR workflows.

Using advanced AI for TAR

While advanced AI can be more effective, and often necessary – particularly for large and/or complex matters – many attorneys remain hesitant to move away from traditional TAR technology. As previously noted, there are a variety of factors behind this hesitancy: familiarity and comfort with the older technology, distrust from lack of technical understanding, fear of the inability to get stakeholders on board, concerns about a learning curve, etc. However, we believe that most of these factors can be overcome by education, transparency, and support.

Part of this education is understanding that the process of using TAR in combination with more advanced AI is not much different from an end-user perspective. TAR 1.0 and TAR 2.0 workflows, wherein subject matter experts or reviewers code documents and the tool learns from those decisions to classify the unreviewed documents, can still be applied with a tool that uses advanced AI. The difference is in the results – as a combination of TAR workflows plus advanced AI can provide results more accurately and quickly with increased flexibility. Again, this is because a tool that uses NLP, deep learning, and machine learning can analyze each document from a variety of different angles, including understanding the context of how language is used rather than simply the definition of one-off words. Thus, for buy-in from other stakeholders, the key is finding the right partner who can not only support the implementation of the AI tool, but can also explain how the technology works, as well as the statistical defensibility of the results it yields.

DATA REUSE

Often companies have hundreds of thousands of previously coded documents sitting unused and dormant, effectively locked in archived or inactive databases. When this data can be leveraged to train a more advanced AI tool, it allows legal teams in the current case to immediately reduce the need for eyes-on review of larger portions of documents. In this scenario, a tool that uses advanced AI can amass a wealth of knowledge by ingesting and analyzing the previous attorney review decisions on each document, as well the metadata, language use, data source type, etc., prior to making the traditional TAR statistical predictions in the current case. The outcome is much more accurate predictions on a much wider variety of classifications than had previously been seen with technology within the ediscovery space. This means that not only are advanced AI tools that can reuse data from past matters better at detecting responsiveness, they can also more accurately detect attorney-client privilege, personal information, key documents, and trade secret information.

For example, when a process is in place that allows for the ingestion and analysis of previous attorney-client privilege decisions, advanced AI tools can not only immediately pinpoint similar or identical documents to those that were coded as privileged in past matters, but can also analyze those past privilege calls from a variety of angles to make much more accurate predictions on privilege on "new" documents.

Applying advanced analytics at a portfolio level can also make prior coding available for reviewers for additional context and to ensure consistency. Coding decisions could also be ported from a past matter to a new one at hand via hash values, depending on the capability of the AI tool itself.

BIG DATA

When datasets are smaller and the data sources within them are homogenous (essentially if a dataset looks like typical datasets from 10 years ago), traditional TAR can still be a particularly useful tool to classify data and reduce eyes-on review. However, when datasets fall within the realm of "big data," an advanced-AI TAR tool is critical to getting accurate classification. Big data is a dataset where the high volume, velocity, veracity, and/or variety makes it difficult to process using traditional tools.[35] As previously noted, the older supervised machine learning utilized by traditional TAR methods evolved before the big data datasets we are seeing in ediscovery today, meaning that technology cannot scale and may experience latency or just general ineffectiveness when faced with this type of scenario. Advanced AI systems like NLP and deep learning not only scale to meet the demands of big data, they also thrive in a big data environment. The more knowledge and data these systems can learn from, the better they are at classifying new data.

Connecting these big datasets for analysis also means gaining a view of an entire portfolio of matters, gaining ediscovery and data insights, and enabling benchmarking.

HSR SECOND REQUESTS

As noted above, TAR (specifically TAR 1.0) has always been especially helpful in HSR Second Requests. And like traditional ediscovery datasets, HSR Second Request datasets are becoming more voluminous and complicated by modern data. Further, HSR Second Requests themselves are also becoming more frequent: February of 2021 marked a 10-year high in HSR activity.[36] The scrutiny on mergers and acquisitions will only continue to increase in the coming years, as President Biden has made antitrust a focal point of his administration – he recently issued an Executive Order to establish a "whole of government" effort to promote competition.[37]

This makes modern HSR Second Requests a perfect use case for a TAR 1.0 process that is enhanced by advanced AI. Negotiations in HSR Second Requests with government regulators around custodians and the relevant time frame for data collection often take up valuable time, making production deadlines much shorter than they would be otherwise. With a TAR 1.0 workflow enhanced by advanced AI, case teams can start analysis and review as soon as the technology and TAR 1.0 workflow is approved by the DOJ or FTC. This is because advanced AI technology can keep the model stable, even as data is in flux – meaning the original training model and classification index will not have to be rebuilt when documents are removed or added from the dataset due to ongoing negotiations. This differs from the legacy technology used in TAR 1.0, where case teams would likely have to start over from scratch when documents are removed or added (wasting valuable time and increasing costs for the underlying organization). In other words, a TAR 1.0 workflow enhanced by advanced AI in a HSR Second Request gives case teams the ability to stay nimble and work much more efficiently.

Additionally, the higher precision gained by advanced AI allows review teams in HSR Second Requests to confidently eliminate larger swaths of nonresponsive documents from production. In other words, a TAR 1.0 workflow enhanced by advanced AI can result in a much smaller responsive set at the end of the process (even if the number of training and control set documents remains roughly similar to a traditional TAR 1.0 workflow without advanced AI). In fact, we have seen advanced AI reduce the number of responsive documents in a HSR Second Request by over 40 percent.

Another key benefit with advanced AI is that responsive and privilege classification can happen concurrently (rather than waiting to do one at a time, a la traditional TAR workflows). As previously noted, advanced AI systems are built for big data, meaning they can quickly scale to meet volume challenges while requiring less human training. The ability of advanced AI to handle large amounts of complicated data within shorter time frames, with less manual review and more accurate analysis, allows case teams to quickly pivot based on government negotiations and thus meet notoriously tight HSR Second Request deadlines.

In fact, within the last year we have seen the government accept this type of technology (a TAR 1.0 workflow enhanced by advanced AI) during HSR Second Request negotiations. In accepting TAR enhanced by advanced AI, it is clear that government regulators are also beginning to understand the power of this technology to simplify and speed up the HSR Second Request process. As such, we expect to see more and more regulators approving advanced AI in HSR Second Request discovery.

Regulator knowledge and understanding around the capabilities of advanced AI can also be a double-edged sword for those unwilling or unable to evolve with technology. Protecting sensitive information like protected health information (PHI) or personally identifiable information (PII) amongst millions of documents under short HSR Second Request timeframes used to be a herculean task. As such, regulators were more willing to allow for mistakes and inadvertent disclosures with little to no penalties or repercussions. However, as more and more states enact stricter data privacy laws while at the same time regulators see how accurately and efficiently advanced AI can pinpoint such data within large datasets, the time is coming when parties will be expected to use that technology to protect personal information from disclosure.

Conclusion

In the long and storied history of the legal profession, ediscovery is still new. But this doesn't mean attorneys and ediscovery practitioners can become complacent with the TAR technology of the last two decades. While in comparison to the history of litigation, a decade can seem like the blink of an eye. But for technology outside of the legal space, a decade is a lifetime.

While the introduction of TAR in the mid-to-late 2000s was revolutionary to the legal industry, analytics leveraging the same basic machine learning technology had been assisting decision making in nonlegal industries since the 1960s.[38] And likewise, the technologies now being introduced to ediscovery have been used for years in other industries. Thus, there are few concerns in stability, and best practices for development and use are well known.

The challenges that modern data poses to traditional ediscovery workflows are substantial and will only continue to grow. To survive and meet those challenges (as well as comply with discovery obligations), it is necessary to evolve *with* technology rather than fight it – just as the legal industry has done in the past with TAR. Advanced AI technology is (and will continue to be) key to this evolution – allowing law firms and case teams to rise to the challenge and manage the increasing volumes and variety of modern datasets of today and tomorrow.

Leaders of corporate legal teams can no longer afford to use technology that is 40 years behind their colleagues. The rise of modern data and use of business intelligence across entire organizations means that it is time to join nonlegal disciplines in adopting new technologies built for big data to maximize budgets, optimize resources, and make strategic business decisions, all backed by data.

References

1 Ragland Jr, George. "Discovery before Trial" Chicago: Callaghan and Company, 1932. https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1015&context=michigan_legal_studies

2 Kershaw, Anne. "Automated Document Review Proves Its Reliability." Digital Discovery and e-Evidence. 5. 2005. http://www.akershaw.com/Documents/2004AEKDocReviewArticle.pdf

3 Id at 1

4 David Bruno. "TAR 2.0 and the Case for More Widespread Use of TAR Workflows." Lighthouse Illumination eDiscovery Blog. February 5, 2021.

5 Gricks, Thomas C. III; Ambrogi, Robert J. "A Brief History of Technology Assited Review." November 17, 2015. https://judicialstudies.duke.edu/wp-content/uploads/2017/07/Panel-1-A-Brief-History-of-Technology-Assisted-Review.pdf

6 Grossman, Maura R.; Cormack, Gordon V. "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review." XVII RICH. J.L. & TECH. 11 (2011). http://jolt.richmond.edu/v17i3/article11.pdf

7 Id.

8 Da Silva Moore v. Publicis Groupe & MSL Group, No. 11 Civ. 1279 (ALC) (AJP) (S.D.N.Y. Feb. 24, 2012) https://scholar.google.com/scholar_case?case=6856971937505165396&q=da+silva+moore+v.+publicis+groupe&hl=en&as_sdt=40000006&as_vis=1

9 Id at 1.

10 Pace, Nicholas; Zakaras, Laura. "The Cost of Producing Electronic Documents in Civil Lawsuits." April 11, 2012. https://www.rand.org/pubs/research_briefs/RB9650.html

11 Bruno, David. "TAR 2.0 and the Case for More Widespread Use of TAR Workflows." Lighthouse Illumination eDiscovery Blog. February 5, 2021.

12 Stevens, Chris. "The Debate is on: TAR 1.0 or TAR 2.0 for HSRs?" Lighthouse Illumination eDiscovery Blog. May 8, 2020.

13 Id.

14 Bruno, David. "TAR 2.0 and the Case for More Widespread Use of TAR Workflows." Lighthouse Illumination eDiscovery Blog. February 5, 2021.

15 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_whatisai.pdf

16 Id.

17 IBM Cloud Education. "Machine Learning." July 15, 2020. https://www.ibm.com/cloud/learn/machine-learning

18 IBM Cloud Education. "Artificial Intelligence (AI)." June 3, 2020. https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

19 IBM Cloud Education. "Deep Learning." May 1, 2020. https://www.ibm.com/cloud/learn/deep-learning

20 IBM Cloud Education. "Natural Language Processing (NLP)." July 2, 2020.
https://www.ibm.com/cloud/learn/natural-language-processing

21 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_whatisai.pdf

22 IBM Cloud Education. "Artificial Intelligence (AI)." June 3, 2020.
https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

23 Id.

24 IBM. "What is big data analytics?"
https://www.ibm.com/analytics/hadoop/big-data-analytics

25 IBM Cloud Education. "Deep Learning." May 1, 2020.
https://www.ibm.com/cloud/learn/deep-learning

26 Herckerman, D. "Bayesian Graphical Models and Networks." Microsoft Corporation. November 2, 2002.
https://www.sciencedirect.com/science/article/pii/B0080430767004411?via%3Dihub

27 IBM Cloud Education. "Machine Learning." July 15, 2020.
https://www.ibm.com/cloud/learn/machine-learning

28 Id.

29 IBM Cloud Education. "Artificial Intelligence (AI)." June 3, 2020.
https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

30 IBM Cloud Education. "Natural Language Processing (NLP)." July 2, 2020.
https://www.ibm.com/cloud/learn/natural-language-processing

31 Josh, Naveen. "7 Types of Artificial Intelligence." Forbes. June 19, 2019.
https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/?sh=39186cb8233e

32 Cross, Rob; Parker, Andrew; Borgatti, Stephen. IBM Institute for Business Value. A bird's-eye view: Using social network analysis to improve knowledge creation and sharing. 2002.
https://www.researchgate.net/profile/Andrew-Parker-28/publication/249858768_A_bird's-eye_view_Using_social_network_analysis_to_improve_knowledge_creation_and_sharing/links/0f31752e6194a6cecf000000/A-birds-eye-view-Using-social-network-analysis-to-improve-knowledge-creation-and-sharing.pdf

33 IBM Cloud Education. "What is Supervised Learning." August 19, 2020.
https://www.ibm.com/cloud/learn/supervised-learning

34 IBM Cloud Education. "What is Unsupervised Learning." September 21, 2020.
https://www.ibm.com/cloud/learn/unsupervised-learning

35 IBM. "What is big data analytics?"
://www.ibm.com/analytics/hadoop/big-data-analytics

36 Maribeth Petrizzi; Heather M. Johnson "HSR Early Termination After a Second Request Issues." March 12, 2021.
https://www.ftc.gov/news-events/blogs/competition-matters/2021/03/hsr-early-termination-after-second-request-issues

37 "FACT SHEET: Executive Order on Promoting Competition in the American Economy." The White House Briefing Room. July 9, 2021.
https://www.whitehouse.gov/briefing-room/statements-releases/2021/07/09/fact-sheet-executive-order-on-promoting-competition-in-the-american-economy/

38 BOLCH Judicial Institute and Duke Law. “Technology Assisted Review (TAR) Guidelines.” January 2019. https://edrm.net/wp-content/uploads/2019/02/TAR-Guidelines-Final.pdf

LIGHTHOUSE