

Unlocking the Benefits of Active Learning

4 Steps to a Successful Implementation

Aaron R. Thorp | IPRO.

The use of Analytics technology has been at the forefront of virtually every industry for years and often supplements the driving of innovation, facilitates decision making, provides insight and analysis, aids in the detection of trends, uncovers patterns, and many other examples.

Although the legal industry often lags other industries in the use and adoption of technology, the use of Analytics is an area that continues to grow and become more mainstream. In fact, the use of Analytics has not only been accepted by courts but is frequently encouraged.



Former U.S. Magistrate Judge Peck, arguably the most notable proponent of Analytics and assisted review, wrote in his landmark 2012 case, *Monique Da Silva Moore, et al. v. Publicis Groupe & MSL Group*, the first judicial decision approving the use of technology-assisted review, and stated, “...assisted review is an acceptable way to search for relevant ESI [Electronically Stored Information] in appropriate cases”. Judge Peck wrote again in his 2015 case, *Rio Tinto PLC v. Valle S. A.*, the decision that it was black-letter law that if the responding party wished to use TAR, courts would allow it.

Technology Assisted Review (“TAR”) is a term often used to define assisted review workflows that involve leveraging analytics. This paper will use the term, “Active Learning”, and explore ways of successfully implementing it.

At the highest and simplest level, Active Learning is the process of taking the document review decisions from human reviewers, building a model from the input, and then classifying documents within the case, based on the model. The system also assigns a relevancy score to each document that can be used to determine how relevant or non-relevant the model predicted it to be. Through this process, teams can quickly prioritize their review efforts and focus on documents predicted as relevant, or ones that have a higher likelihood of being relevant. Similarly, teams can spend less (or even no) time on documents predicted as non-relevant, or ones that have a higher likelihood of being non-relevant. Although Active Learning workflows are most often associated with identifying relevant documents, workflows can also be implemented to find documents that are hot, related to specific issues, or to assist in privilege review. (1)

Active Learning has been shown to be more accurate than manual review efforts, with significantly less effort. “The myth that exhaustive manual review is the most effective – and therefore, the most defensible – approach to document review is strongly refuted” (2). By implementing Active Learning workflows, teams can drastically reduce the time and costs of document review, provided one follows a reasonable and defensible workflow.

Key Terminology

Before we get too far in the discussion, it is important to discuss some basic terminology. This is certainly not a comprehensive list, but when combined with the additional terms and definitions in the Glossary, should serve as a good baseline for implementing Active Learning. These key terms are frequently encountered, not only within legal analytics and Active Learning workflows, but used throughout the field of statistics and data science.

Prevalence, also referred to as Richness, is the percentage of relevant documents in your case. In other words, if we have a case of 1 million documents, and 100,000 are relevant, our prevalence would be 10% ($100,000 / 1,000,000$). Prevalence can vary dramatically across cases and is influenced by a number of factors, such as the type of case, the industry or business unit, and even the collection strategy (i.e., was it a very targeted collection with likely higher prevalence, or was it a broader collection, with likely lower prevalence). The important point to keep in mind is that the lower your prevalence, the more difficult it will be to find relevant documents, typically leading to more time, effort, and cost.

Precision is the percentage of documents identified and returned in a search or review effort, that actually are relevant, sometimes described as, “a measure of accuracy”. In other words, if the system classified 10,000 documents as relevant, but only 9,000 were truly relevant (and 1,000 were truly non-relevant), our precision would be 90% (the formula for precision is: $TP / (TP + FP)$, or $9,000 / (9,000 + 1,000)$). High precision is desirable for review cost purposes, because it means the model has done a good job of identifying relevant documents, meaning our review team is not wasting time reviewing non-relevant documents the model incorrectly thinks are relevant. If our model had 50% precision, documents predicted as relevant would have about a 50/50 chance of being truly relevant, meaning our reviewers would not only be frustrated, but we would be spending a lot more time, cost, and effort as compared to a model with higher precision.

Recall is the percentage of the total amount of relevant documents that are retrieved by a search or review effort, sometimes described as, “a measure of completeness”. In other words, if we have a case of 1 million documents, we’ll assume a prevalence of 10%, so 100,000 documents are truly relevant. Furthermore, our model locates 70,000 documents, that truly are relevant. However, the model “missed” 30,000 documents it identified as non-relevant, that are truly relevant. Our recall would be 70.0% (the formula for recall is: $TP / (TP + FN)$, or $70,000 / (70,000 + 30,000)$). High recall is desirable for defensibility purposes, because it estimates whether we found a reasonable number of relevant documents in our case. Recall is typically the metric legal teams want to focus on the most, because it essentially identifies if our review effort is finding enough

relevant documents to meet our legal obligations. Currently, the industry tends to consider 75% - 80% Recall as reasonable, and certainly the higher, the better. Again, keep in mind, the prevalence of relevant documents in your case will influence your recall.

F1 is the harmonic mean of Precision and Recall, used in order to measure the effectiveness of a search or retrieval process. It is worth pointing out that although Precision and Recall can both be high, it is almost inevitable that as you try to increase Recall, your Precision will go down, as there is typically a tradeoff with precision and recall. A very important point to keep in mind is that the F1 score gives equal weight to both Precision and Recall, which may not be desirable for document review purposes. Remember that Precision is important for our internal review team's cost purposes, but it does not indicate whether we found all, or even a sufficient number of relevant documents. On the other hand, Recall is the metric that indicates whether we have found enough relevant documents in order to meet our legal obligation. Consequently, Recall is often the most important metric to track, and will occasionally even be negotiated upfront with opposing parties, but it is really best to have a firm grasp of your case, understand your data, and be comfortable with Active Learning metrics and workflows, before you consider agreeing at the onset to specific Recall levels. The merits of a case are really the driving factor as to whether Precision or Recall is more important. Take for example, a case that has very little monetary value at stake. It might make sense to really focus efforts on Precision, in order to keep review costs down, and achieve just enough Recall in order to meet the level of reasonableness required by law. In that scenario, it really is not worth the extra effort and cost to find additional relevant documents. On the other hand,

consider a, "bet the farm" kind of case with significant risk at stake. In that situation, it might not matter how much cost and effort is spent on review, because we are really most concerned with Recall, and locating as many relevant documents as possible. The main takeaway is that the F1 score arbitrarily weighs precision and recall equally, without taking into consideration other variables of your case or strategy. As one industry expert pointed out, "it is better to compare algorithm performance by comparing precision at the same level of recall, with the recall chosen to be reasonable for a case." (4)

Understanding Active Learning with Ipro Review

Let's discuss at a high level, the Ipro approach to Active Learning. By design, Active Learning is enabled by default. When documents are loaded in a case, text from your documents is first analyzed. The process identifies whether each document has enough, good quality text in order to make relevancy predictions. Consider for instance, the email message that says something along the lines of, "See attached", or even the graphics/images that might not have any text whatsoever. There is simply not enough text in those particular examples to predict whether it would be relevant or non-relevant, so those types of "short" documents are set aside. Similarly, we might have documents such as spreadsheets that might have more text, but it could be completely numeric in nature. Those "binary" documents would also be set aside, as it would be difficult, if not impossible, to identify a document as relevant or non-relevant with purely numeric values.

What we're left with, is typically the vast majority of documents in your case, containing enough text that allow the algorithm to extract significant words from each document. These significant words then allow us to group documents together into potentially hundreds or thousands of very small "clusters" of

conceptually similar documents. A very simple analogy might be the process you go through when putting together a jigsaw puzzle. When you first start, you might group together pieces by sizes, shapes, colors, patterns, borders vs no borders, corners, dogs, cats, birds, etc., so you end up with a number of piles (or clusters) containing similar pieces. A fairly similar process happens behind the scenes when starting an Active Learning project.

After the creation of your review pass, and the identification of the tag you plan to use for tagging a relevant document, the Active Learning process begins. At the onset of review, documents are strategically pulled from the clusters we discussed earlier, and presented to the reviewer for tagging. If we revisit the jigsaw puzzle example again, perhaps I am looking for a puzzle piece containing a tan colored puppy, with 4 edges, and also containing a border. Someone else could show me a single piece from each pile (or cluster), and then I could tell them if it is similar (or relevant) to what I am interested in. This is a simplistic example but is somewhat similar to how documents are grouped into clusters and then strategic documents are presented to the reviewer for review. The reason this is an efficient approach is because if I am presented a puzzle piece from the pile containing black cats without any borders, then I can quickly skip over and avoid looking at other puzzle pieces from the same pile, thereby saving significant time, cost, and effort. This type of assisted review workflow is known as TAR 3.0. Bill Dimm was the first to coin the term TAR 3.0 in 2015 and has written and spoke about how it performed better than other assisted review workflows. (5)

Returning to our document review workflow, as reviewers are tagging and training the system with more and more examples of relevant vs non-relevant documents, the system continues to build and refine

its algorithm. Reviewers are continually presented with documents predicted as relevant, in a prioritized manner, allowing reviewers to review and tag documents much quicker. As the review continues, relevant documents are prioritized and ordered by the highest relevancy ranking. Once we reach a point of exhausting the highly ranked cluster examples, the algorithm will shift and start presenting any document that is predicted as relevant. Over time, the trend will continue until relevant documents start tapering off and decrease, only to be replaced by non-relevant documents. Eventually, we will be encountering mostly non-relevant documents, perhaps only finding a relevant doc after reviewing hundreds or more non-relevant documents. The cost to find relevant documents will continue to increase as we need to review and go through more and more non-relevant documents, with the goal of increasing our Recall, but as a side effect, will be decreasing our Precision.

Current Approach

Historically, the most straightforward approach to reviewing documents was to simply add reviewers to a review team, assign them batches of documents, and review everything in the database.

Sure, there may have been some combination of date ranges, keywords, analytics, or other strategies used upfront to either cull the data or prioritize it for specific reviewers or subject matter experts, but it was still essentially the same manually intensive review workflow that focused on throwing bodies at the problem. Keyword searching was (and often still is) the historic go-to approach for identifying relevant documents but studies have even shown how Active Learning and other assisted review workflows outperform keyword searching. (6)

Although we continue to see the increased adoption of Active Learning workflows, indeed, many firms still continue to review documents using this dated approach to document review today, for various reasons. Let us now discuss several steps that can modernize outdated, labor intensive workflows, by implementing Active Learning. I like to think of the steps using the Crawl, Walk, Run, Fly methodology.

Step 1: Crawl

At the most basic level, implementing Active Learning on a case, will increase review speed and efficiency.

For those just starting to employ Active Learning, the first huge benefit will be the way that documents are presented and served up to reviewers. At the beginning, you will tag documents until you encounter a relevant document. At that point, documents are then prioritized using the strategic cluster approach discussed earlier. Essentially, it means you will be reviewing documents in a more efficient approach and not wasting time reviewing documents that may be conceptually similar to each other. For example, let's say we have 10 documents in our case that are all conceptually similar to each other and part of the same cluster (if you recall from earlier, they were grouped together in the same cluster based on the significant words associated with the 10 documents). We'll assume the documents are all Vacation Request Forms from our client's documents. We'll further assume these documents are not relevant to our case. During the initial phase of our review, only 1 document is presented to reviewers for tagging as either relevant or non-relevant (non-relevant in our example here). That means our reviewers do not need to spend time reviewing the other 9 similar documents but can instead review other, more diverse examples from the case. Throughout this process, reviewers are not only tagging documents but training the system at

the same time as they go along. Review and training are happening simultaneously.

This review approach continues, until highly ranked, predicted relevant cluster documents are exhausted and then any predicted relevant documents in the case are presented to reviewers, prioritized by relevancy. If a team prefers to review every document in the case, you would simply continue to review all documents. Over time, the relevant documents will typically taper off and be replaced with non-relevant documents. This means a team can plan their review and perhaps strategize when and who does the review. It might make sense to have your more experienced (but more costly) reviewers do much of the tagging upfront, and then over time, they could be supplemented or even phased out, to be replaced with less experienced (but less costly reviewers). Many of the documents at the end of a review will tend to be non-relevant (and predicted as such), so it only makes sense to consider alternative workflows.

The merits of the case, reasonableness and proportionality, will dictate the workflow approach, and it is certainly possible you could run into a group of relevant documents towards the end that are unlike any others encountered previously, especially if your understanding of relevancy shifts over the course of your case. However, the key takeaway is that although you are still reviewing all documents in your case (similar to how you might do it today), you are reviewing in a much more efficient manner.

Step 2: Walk

Another approach to Active Learning is to determine a reasonable stopping point in the review process, leaving some unreviewed, but predicted non-relevant documents behind.

Similar to the first step, we implement Active Learning on our case, but as we notice relevant documents taper off, we may decide to stop reviewing. A great analogy used to describe the tapering off process of relevant documents over time, is when you make popcorn in the microwave. In the beginning, there is some activity as the popcorn kernels heat up, gradually progressing into a frenzy of popcorn, repeatedly transforming from kernels to fully popped, and burst popcorn. Then, activity starts to gradually decrease as the vast majority of popcorn has already been popped. Eventually, we notice there is very little, if any popcorn action. Sure, we might get a few stragglers and outliers here and there as things slow down, but essentially our popcorn is finished and we're ready to take it out and enjoy it. When we open the bag, we expect to see only popcorn but it is not surprising to find a handful of unpopped kernels. Overall, the vast majority of our popcorn making exercise was successful, even if we pulled the bag out a little early, and were left with a few of those remaining kernels. We knew when to pull the bag out of the microwave because we listened to the tapering off and felt it was reasonable to discontinue more cooking. (7)

The Active Learning workflow for relevant documents is very similar to our popcorn example. As our relevant documents diminish, we can determine a reasonable stopping point by tracking metrics such as our Recall estimate. Many courts and industry experts use something in the neighborhood of 75% - 80% Recall as reasonable. Although at first glance, this means we are accepting the fact that we missed 20% - 25% of relevant documents, it has been proven that Active Learning is better than manual, human review, and has been widely accepted by the courts. It is important to keep in mind the goal of document review is not absolute perfection. Indeed, studies show even entirely human review efforts are not

tagging documents with absolute perfection or locating every single relevant document within a case, but instead meet a "reasonable" expectation. (8)

One recent study pointed out that although you may only be locating 80% of relevant documents, you are missing very few, if any, relevant facts that have not already been identified in other relevant documents. The idea is that a case is really about relevant facts, or topics, and documents typically contain one or more topics, that frequently overlap and exist in many other documents. The study goes on to state, "we find no evidence that there is significant value in continuing to search for new information once a moderate level of Recall has been achieved (say around 80%)". (9)

Step 3: Run

We determined it is reasonable to suspend review after achieving a reasonable Recall, but in order to provide more defensibility, we will want to validate our results through sampling and additional QC processes.

As we discussed in the previous step, once you achieve your desired Recall, it is often considered acceptable to stop reviewing for additional relevant documents. However, we are still left with many (potentially thousands), of predicted non-relevant, but still unreviewed documents, that a human reviewer has not put eyes on.

This is the point when teams typically engage in any number of QC workflows, in order to demonstrate that a reasonable number of relevant documents have been found, and continuing to search for more documents is not worth the cost and effort, because we have reached the point of diminishing returns.

One approach might be to calculate the cost to review each document. Teams will often have detailed metrics on the average cost to review

documents and will sometimes break it out even further to calculate the cost based on the type of reviewer activity (i.e, first pass review, privilege review, applying redactions, etc.). In our case, if we estimate our cost to review each document at \$2.00 per document, we can quickly start determining how much more it will cost to continue reviewing. Perhaps, we have reached that point in our review of diminishing returns, when we only encounter a relevant document after reviewing 1,000 non-relevant documents. If we have 25,000 predicted non-relevant (but unreviewed) documents remaining, with some quick math, we estimate it would cost us another \$50,000 (25,000 documents X \$2.00 per document = \$50,000) to continue our review. Furthermore, that might only turn up another 25 relevant documents (based on our calculation of 1 relevant doc for every 1,000 non-relevant doc ratio). Is it worth the time, cost, and effort to locate another 25 relevant documents, if it costs us \$50,000? Variables such as reasonableness, proportionality, timing, case strategy, merits of the case, risk tolerance, and certainly other factors will all play into the decision, but it is certainly something teams should be considering.

Document conflicts are another particularly key area for QC. Conflicts, or discrepancies, arise when the Active Learning algorithm predicted a document one way, but a reviewer tagged it the other way. Conflicts are certainly bound to happen, especially as the algorithm is receiving more and more training input and refining its predictions. Fortunately, it is easy to identify conflicts, as the system will identify them and allow you to quickly target them for further scrutiny. Teams will often review conflicts, especially if the score strongly indicates it differently than how a human reviewer tagged it. It could be a simple explanation such as the reviewer just inadvertently tagged it the wrong way. Case teams could review all conflicts, or simply focus on the ones with scores

that differ significantly from how the reviewer tagged the document. The QC of conflicts offer case teams another method to identify and minimize those types of errors.

Once you have reached your diminishing returns, found most relevant documents, and achieved a reasonable level of recall, you may want to perform further analysis on the remaining documents through an elusion test. An elusion test is a type of validation process that samples from your discard pile - the predicted non-relevant documents you plan to leave behind without review or production. Ideally, we prefer not to find any relevant documents when sampling the predicted non-relevant documents, but as discussed before, the goal of review is not perfection. There are numerous factors that will determine what is considered a reasonable percentage of relevant documents in our discard pile, but we certainly want to keep the percentage at an absolute minimum. For example, if we found that 50% of documents in our discard pile were in fact relevant, then we would need to continue reviewing and training the system, as it would be unlikely that any opposing party or court would find that as acceptable, or a reasonable effort. If, on the other hand, the percentage of relevant documents in our discard pile was around 2%, it is very possible one could make an argument we may be at a suitable stopping point. Keep in mind, the prevalence of our case will play a significant role, not only in the percentage of relevant documents deemed acceptable, but will also determine an appropriate sample size for the elusion test. (10) An elusion test is a valuable tool, especially combined with other validation and QC methods.

We also should not forget about QC workflows for manual, human reviewed and tagged documents. These might be somewhat similar to what teams employ today, even if they are not already using

Active Learning, since it is performing some type of quality control on reviewer decisions. Senior reviewers may be tasked with sampling and checking the work of junior reviewers. You could also perform a daily random sample of a small percentage of tagged relevant or non-relevant documents, which can be tracked over time, to calculate the daily overturn rate (documents that were tagged one way, but were later changed for one reason or another). During this process, it would also be possible to identify specific reviewers that are more prone to tagging documents incorrectly, or have more disagreements with the relevancy predictions, leading to more conflicts. There could be any number of next steps at that point, but you have the information to make an informed decision.

Step 4: Fly, and Beyond

The last step in our journey moves on to more advanced workflows and scenarios. As case teams become increasingly comfortable using Active Learning for relevancy review and prioritization, the next logical question is whether it can be used for privilege review. Because privilege review is so unique, and is such an important step in the review workflow that requires great care and attention, it is still often done by a human review team. Privilege review often involves the scrutiny of subtle textual or metadata clues, analyzing recipients to an email, or other factors. However, that is not to say an Active Learning workflow could not be utilized as a QC step to supplement the human review effort for privilege documents. In fact, there is virtually no downside to using Active Learning for any QC workflow, when paired with human reviewers as an additional layer of protection.

Similar to the review of privilege documents, the review for confidentiality is still often performed by a human review team. When there are multiple levels

of confidentiality within a case, it would be difficult for Active Learning to distinguish between a document that is confidential and then another that may be highly confidential. The use of Active Learning to assist in identification of confidential documents, could be a great QC workflow addition that follows human review, just as before with the privilege review.

As more and more cases span the globe, they often include foreign language documents, potentially bringing additional challenges to Active Learning workflows. It is not a requirement, but when faced with a case that has multiple languages, it may be beneficial to divide documents by language and setup a Active Learning workflow for each language. The model uses an algorithm to make predictions, so it does not “read” or “understand” the specific language, but it will often perform better when the model is specific to one language.

After successfully implementing an Active Learning workflow, often one of the final remaining steps is to handle items within a document family. The workflow may vary from matter to matter but you will often be left with many documents that have conflicting relevancy calls within the family. In other words, the email might be considered non-relevant, but the attachment considered as relevant. Just as with a purely human review effort, there is often a workflow implemented to analyze document families with conflicting predictions and then tagged accordingly for final production.

Other Considerations

Below are a few final questions and considerations that may not have easy answers, but are some key takeaways and action items to consider when implementing Active Learning workflows.

Has the legal team determined the type of Active Learning workflow(s) that will be implemented on the case (i.e., full review of all documents, limited review, QC review, etc.)?

Have documents already been culled prior to implementing Active Learning?

Are there more documents that need to be processed and loaded?

Is there any existing work product or already reviewed documents that can be supplied for Active Learning training at the beginning of a case?

Did the opposing party and/or court require or agree on a certain level of Precision, Recall, acceptable elusion rate, or other measures?

Did the review team account for additional time and review to handle documents that are not candidates for Active Learning (e.g., images, audio/video, documents with a small amount of text, spreadsheets, etc.)

Are there subject matter experts that might be more valuable upfront to get a good start on the review/training?

If you are not planning to review all documents in your case, do you have a well thought out QC plan and feel comfortable explaining your workflows, especially if you cannot achieve your desired level of recall?

Conclusion

Incorporating Active Learning into your review workflows is much easier than it seems and offers tremendous benefits. Active Learning typically reduces review time substantially, saving on costs and review efforts, while increasing overall review performance. After moving to Active Learning workflows, teams often find they cannot go back to the old way of review and often wonder how they managed document review before. Even smaller cases benefit from the prioritization of documents and getting review teams quickly focused on the most relevant documents in a case. The use of Active Learning can also be a great way to level the playing field for smaller firms, in order to compete against large firms. Ipro makes it even easier to incorporate Active Learning into workflows, as it is “always on” and running in the background, even if you choose not to use it at the onset of a case.

Glossary

Terms are quoted directly from the Grossmann-Cormack glossary of Technology Assisted Review, and grouped together with similar terms. (3)

- **Prevalence:** The fraction of Documents in a Population that are Relevant to an Information Need. Also referred to as Richness or Yield.
- **Precision:** The fraction of Documents identified as Relevant by a search or review effort, that are in fact Relevant. Also referred to as Positive Predictive Value.
- **Recall:** The fraction of Relevant Documents that are identified as Relevant by a search or review effort.
- **F1:** The Harmonic Mean of Recall and Precision, often used in Information Retrieval studies as a measure of the effectiveness of a search or review effort, which accounts for the tradeoff between Recall and Precision. In order to achieve a high F1 score, a search or review effort must achieve both high Recall and high Precision.
- **Precision-Recall Curve:** The curve representing the tradeoff between Precision and Recall for a given search or review effort, depending on the chosen Cutoff value. See Recall-Precision Curve.
- **Precision-Recall Tradeoff:** The notion that most search strategies can be adjusted to increase Precision at the expense of Recall, or vice versa. At one extreme, 100% Recall could be achieved by a search that returned the entire Document Population, but Precision would be low (equal to Prevalence). At the other extreme, 100% Precision could be achieved by a search that returned a single Relevant Document, but Recall would be low (equal to $1/N$, where N is the number of Relevant Documents in the Document Population). More generally, a broader search returning many Documents will have higher Recall and lower Precision, while a narrower search returning fewer Documents will have lower Recall and higher Precision. A Precision-Recall Curve illustrates the

Precision-Recall Tradeoff for a particular search method.

- **Elusion:** The fraction of Documents identified as Non-Relevant by a search or review effort that are in fact Relevant. Elusion is estimated by taking a Random Sample from the Null Set and determining how many (or what Proportion of) Documents are actually Relevant. A low Elusion value has commonly been advanced as evidence of an effective search or review effort (see, e.g., Kleen), but that can be misleading because it quantifies only those Relevant Documents that have been missed by the search or review effort; it does not quantify the Relevant Documents found by the search or review effort (i.e., Recall). Consider, for example, a Document Population containing one million Documents, of which ten thousand (or 1%) are Relevant. A search or review effort that returned 1,000 Documents, none of which were Relevant, would have 1.001% Elusion, belying the failure of the search. $\text{Elusion} = 100\% - \text{Negative Predictive Value}$.
- **True Positive (TP):** A Relevant Document that is correctly identified as Relevant by a search or review effort.
- **True Negative (TN):** A Non-Relevant Document that is correctly identified as Non-Relevant by a search or review effort.
- **False Positive (FP):** A Non-Relevant Document that is incorrectly identified as Relevant by a search or review effort.
- **False Negative (FN):** A Relevant Document that is missed (i.e., incorrectly identified as Non-Relevant) by a search or review effort. Also known as a Miss.

References

1. Quartararo, et al, Technology Assisted Review Guidelines, Bolch Judicial Institute Duke Law, January 2019 (2019). Available at: <https://judicialstudies.duke.edu/wp-content/uploads/2019/02/TAR-Guidelines-Final-1.pdf>
2. Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, 17Rich. J.L. & Tech11 (2011). Available at: <http://scholarship.richmond.edu/jolt/vol17/iss3/5>
3. Maura R. Grossman & Gordon V. Cormack, The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola, U.S. Magistrate Judge, 7 FED. CTS. L. REV. 1, 32 (2013). Available at: <https://www.fclr.org/fclr/articles/html/2010/grossman.pdf>
4. Bill Dimm, Predictive Coding Performance and the Silly F1 Score, May 8, 2013 (2013). Available at: <https://blog.cluster-text.com/2013/05/08/predictive-coding-performance-and-the-silly-f1-score/>
5. Bill Dimm, TAR 3.0 Performance, Jan 28, 2016. Available at: <https://blog.cluster-text.com/2016/01/28/tar-3-0-performance/>
6. Bill Dimm, TAR vs. Keyword Search Challenge, May 8, 2018 (2018). Available at: <https://blog.cluster-text.com/2018/05/08/tar-vs-keyword-search-challenge/>
7. Maura R. Grossman & Gordon V. Cormack, Continuous Active Learning for TAR, April/May 2016 (2016). Available at: https://plg.uwaterloo.ca/~gvcormac/caldemo/AprMay16_EdiscoveryBulletin.pdf
8. The Sedona Conference, TAR Case Law Primer, 18 Sedona Conf. J. 1 (2017). Available at: https://thesedonaconference.org/sites/default/files/publications/TSC%20TAR%20Case%20Law%20Primer%202017_1.pdf
9. Herbert L. Roitblat, Is there something I'm missing? Topic Modeling in eDiscovery, Jul 30, 2020 (2020). Available at: <https://arxiv.org/abs/2007.15731>
10. Lilith Bat-Leah, There Is No One Size Fits All Sample Size Appropriate for TAR Validation (Part II), March 12, 2020 (2020). Available at: <https://www.jdsupra.com/legalnews/there-is-no-one-size-fits-all-sample-42659/>

About the Author

Aaron R. Thorp is a Client Success Manager within the Professional Services Group at IPRO. He has over 20 years of eDiscovery and litigation support experience from both the law firm and service provider sides of the industry. His expertise includes advising clients and developing processes, procedures, and best practices for the use and implementation of technology related to eDiscovery.

Copyright © IPRO TECH LLC 2021

All rights reserved. All other brands and product names are trademarks or registered trademarks of their respective owners. This document is neither designed nor intended to provide legal or other professional advice but is intended to be a starting point for research about Active Learning and electronic discovery. While every attempt has been made to ensure the accuracy of the information in this document, no responsibility can be accepted for errors or omissions.