TAR 1 Reference Model: An Established Framework Unifying Traditional and GenAI Approaches to Technology-Assisted Review

Tara Emery, et al.

25 Sedona Conf. J. 109 (2024)

The Sedona Conference Journal

Volume 25 Forthcoming 2024

TAR 1 Reference Model: An Established Framework Unifying Traditional and GenAI Approaches to Technology-Assisted Review

Tara Emory, Jeremy Pickens & Wilzette Louis

March 2024

Recommended Citation:

Tara Emory, Jeremy Pickens & Wilzette Louis, *TAR 1 Reference Model: An Established Framework Unifying Traditional and GenAI Approaches to Technology-Assisted Review*, 25 SEDONA CONF. J. 109 (forthcoming 2024), https://thesedonaconference.org/publication/TAR_Reference_Model_Unifying_Traditional_and_GenAI_Approaches.

TAR 1 Reference Model: An Established Framework Unifying Traditional and GenAI Approaches to Technology-Assisted Review

Tara Emory, Jeremy Pickens, and Wilzette Louis[1]

Table of Contents

1. At Redgrave Data, Tara Emory is Senior Vice President of Legal AI Strategy and General Counsel; Jeremy Pickens is Head of Applied Science; and Wilzette Louis is Director of Client Solutions. Thanks to our colleagues Dave Lewis, Mike Kearney, France Jaffe, and Rees Crosby, and to Nick Snavely of Redgrave LLP, for their contributions to this article.

A. INTRODUCTION

In eDiscovery, Generative Artificial Intelligence (GenAI) in the form of Large Language Models (LLMs) may offer more efficient approaches for many tasks, including document review. GenAI algorithms can be used similarly to traditional machine-learning algorithms for this purpose, through a process involving iterative training, sampling, and statistics.

In large matters, different aspects of document review are often divided into different workflows and teams. These workflows often begin with a "first-pass review" in which documents are tagged so they can be easily managed into other workflows (e.g., production or substantive review). Many workflow options exist in which human review teams handle the tagging of documents. When that human effort is alternatively replaced with machine tagging, it is accomplished through a form of Technology-Assisted Review (TAR) known as TAR 1, a workflow that involves tagging documents through the use of predictive algorithms. TAR 1 is applicable to GenAI review when it is used for first-pass review.

GenAI is promising as a new solution and will be a useful approach if it proves to be at least as effective as the options of human review or traditional TAR 1, and comparable in time and costs required. In addition, GenAI's potential to also perform other tasks that are traditionally done after first-level review, such as privilege review or summarization, may further save on costs and time.

The steps of TAR 1 (referred to interchangeably as both a process and a workflow) involve building a predictive model and then demonstrating its effectiveness. These steps ensure that the practitioners who use it are successful in their goals and confident in the outcome. GenAI can be used for first-level review in place of the discriminative machine-learning algorithms that have traditionally been used in TAR 1. Therefore, other

than substituting a new algorithm, the conceptual steps of a TAR 1 process are essentially identical, regardless of which type of predictive algorithm is used. In order to facilitate successful outcomes, GenAI as a predictive algorithm needs to be wrapped in a process known to be familiar, reasonable, effective, and defensible by practitioners doing machine tagging for first-pass review: TAR 1.

We provide a reference model to serve as a foundation for first-pass workflows that use artificial intelligence/machine learning to integrate them into the established process of TAR 1. We also provide diagrams of the tasks within the steps of the reference model for discriminative TAR 1 (TAR 1 using discriminative algorithms) and GenAI TAR 1 (TAR 1 using generative AI algorithms) to demonstrate their similarities and differences. One can view predictive algorithms as engines, while the TAR 1 process is a vehicle. The engines may vary, but the steering, seating, wheels, and other key features of the vehicle are unchanged. To understand what engine to use for different goals, empirical studies are needed on comparative benefits in terms of time, cost, effectiveness, consistency, and other metrics of interest. The TAR 1 reference model may guide those studies, and help practitioners understand the similarities and differences between TAR 1 workflows using traditional discriminative algorithms and those using GenAI.

B. TAR AND GENAI

1. TAR

Coined by Maura Grossman and Gordon Cormack, TAR was defined in the Grossman-Cormack Glossary of Technology-Assisted Review:[2]

> **Technology-Assisted Review (TAR)**: A process for Prioritizing or Coding a Collection of Documents using a computerized system that harnesses human judgments of one or more Subject Matter Expert(s) on a smaller set of Documents and then extrapolates those judgments to the remaining Document Collection. Some TAR methods use Machine Learning Algorithms to distinguish Relevant from Non-Relevant Documents, based on Training Examples Coded as Relevant or Non-Relevant by the Subject Matter Experts(s), while other TAR methods derive systematic Rules that emulate the expert(s)' decision-making process. TAR processes generally incorporate Statistical Models and/or Sampling techniques to guide the process and to measure overall system effectiveness.

As made clear in this definition, TAR is a process that (a) uses subject-matter experts to (b) train a computerized system (algorithm) to make predictions and (c) guides both the training and the results of that process via sampling and various kinds of statistics.[3]

---

2. Maura R. Grossman and Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, 2013 FED. CTS. L. REV. 7 (January 2013).

3. Note also that not all prediction engines are discriminative supervised machine learning, or even supervised machine learning, as articulated in the TAR definition. For example, expert systems are not supervised machine

Most currently available TAR implementations utilize supervised machine learning, more specifically discriminative supervised machine learning, as the prediction engine. A supervised machine-learning algorithm takes human-labeled data (e.g., documents that the human has coded responsive or not responsive[4]) as input, and the machine learns a function that makes predictions on untagged documents. The term "discriminative" refers to a specific kind of predictive algorithm that separates, or discriminates, between positive and negative labels. Common examples of discriminative algorithms used in TAR include support vector machine classifiers and logistic regression classifiers. A myriad of TAR workflows exist, and parties may use different ones to meet different goals. While variations and hybrid approaches exist, TAR 1[5] and TAR 2[6] often describe the most common general approaches.

---

learning in that they do not learn a predictive function from data. Rather, humans derive If-Then rules for the machine to follow, with rules mimicking the types of decisions that an expert would make. A typical eDiscovery workflow involves humans examining documents from a project to learn about their contents, then writing and refining rules, and then testing and measuring the rules' performance. Once a set of rules is fixed, the machine then applies the rules to extrapolate predictions onto untagged documents. This expert systems workflow bears much similarity to the way GenAI is used in TAR, particularly with how the efforts of the human and the machine are divided.

4. While responsiveness (to document requests in discovery) and relevance (to a matter or topic) differ in meaning, they are often used interchangeably in eDiscovery discussions, and our use of one or the other is not intended to be significant in this article.

5. TAR 1 is also known as SAL (Simple Active Learning) and SPL (Simple Passive Learning), and as "two-stage TAR," wherein training and review are two separate activities. Humans iteratively train a model for a finite number of steps, and then the model labels the remainder of the documents.

6. TAR 2 is also known as CAL (Continuous Active Learning™), as well as "one-stage TAR," wherein training and review are the same activity. In

2. GenAI

GenAI, in the form of Large Language Models, is based on deep learning models that have been trained on enormous amounts of text from which they have learned how to predict the next words in a given sequence based on a "prompt." They do not discriminate between classes; rather, they sequentially generate words probabilistically. By itself, an LLM is not a supervised machine learning model; it only does next-word prediction. Suppose, however, that the LLM is fed with the following word sequence:

> I am looking for information about cows. Here is the text of a document: "The farmer went to his barn to put out square bales for the bovines." Is this text relevant to my information need? Please answer yes or no.

The combination of the prompt with the LLM in essence becomes a (generative) supervised machine learning classifier.[7] The instructions to the LLM about the nature of information being sought (cows), combined with the document for which a prediction is desired, plus instructions about the text to generate, form the supervision. The LLM generates text that serves as a prediction. The length of the generated text may be short, but the LLM is nonetheless generative rather than discriminative. While it should most of the time respond with "yes" or "no" as instructed, it selects its response from limitless options, so there

---

TAR 2, humans tag the documents, and the model is updated continuously to include the new examples they have tagged.

7. Specifically, this kind of supervised machine learning is known as "zero shot learning," because the LLM can make predictions about a class of interest (responsive or not responsive to an issue) from a straightforward natural language description of that issue, rather than from labeled training documents.

is also some probability of generating other responses, e.g., "giraffe." All instances of "yes" can be considered responsive and "no" can be considered not responsive. All instances of anything else can either be considered either not responsive or a non-answer (failure to predict) that indicates a need for further review.[8] In this manner, the LLM becomes capable of extrapolating onto untagged documents based on a human-written description of relevant information.

In eDiscovery, GenAI may additionally be used in many other ways, including but not limited to summarizing documents, answering questions, giving explanations, extracting key information, identifying personal information, identifying and reviewing foreign language documents, and privilege logging. This article discusses using GenAI for document review in eDiscovery, specifically, its capability to assist in first-pass review by tagging documents based on its predictions.

3. TAR 1

This article analyzes the TAR 1 workflow to provide a unified lens through which to understand its application regarding different kinds of predictive engines. TAR 1 is a form of TAR in which the machine predictively tags the documents in the project population in two sequential stages: build the model, then classify the population. Its most frequent application is facilitating the selection of documents that are most likely responsive for compliance with production requirements. It generally aims for efficiency by reducing the amount of effort needed to build the predictive model while maximizing the effectiveness (precision and recall) of the predictions. It involves building a model

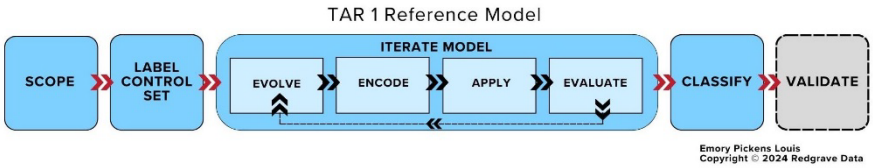
8. The LLM can be asked to generate rankable categories as well, such as "super tippy top relevant," "highly relevant," "relevant," "not so relevant," and "not even close," with cutoffs drawn at different points. There are many possibilities, but the principle remains the same.

to classify the project document population, and then tagging that population (e.g., responsive or not responsive). The project document population, with predicted tags applied, can then be filtered for other tasks, such as selecting documents predicted responsive for production. Therefore, TAR 1 is usually cost-effective and quick compared to workflow options that require human review to tag documents.

While largely bypassing human first-pass review with TAR 1 can greatly save on costs and time, at least for its most common purpose of production compliance, it also introduces certain risks. It will inevitably predictively tag as responsive documents that are not, potentially risking unnecessarily revealing some confidential or sensitive information in productions, and even may risk possible challenges by other parties for overproduction. While complementary workflows are carried out to search for, review, and withhold or redact certain documents (e.g., privileged or personal information), some documents containing such information may be missed if they do not contain the criteria used to create those workflows, like keywords. Therefore, TAR 1-based productions can also risk the inclusion of such information. Furthermore, whereas first-pass human reviewers can tag documents and often identify and communicate insights about the documents for the case team, TAR 1 only tags documents.

C. TAR 1 REFERENCE MODEL

The TAR 1 Reference Model depicts the established, defensible TAR 1 process, in which the effectiveness of the result is measured through sampling and statistics. These five steps of 1) **Scope**, 2) **Label Control Set**, 3) **Iterate Model**, 4) **Classify**, and 5) **Validate** apply regardless of the algorithm used to predict a responsiveness tag, whether that is the traditionally used discriminative algorithms, GenAI, expert systems, or any number of other predictive techniques.



Reprinted with permission from Redgrave Data. Click on graphic for expanded view.

The steps of TAR 1 are:

1. **Scope:** Assemble the project document population and establish the definition of responsiveness
2. **Label Control Set:** Tag a random document sample to estimate the effectiveness of model predictions
3. **Iterate Model:** Create and improve a model to predict responsiveness
    a. **Evolve** the selection of information that will be used to improve prediction
    b. **Encode** the improved information into the model
    c. **Apply** the model to the control set
    d. **Evaluate** the model's performance on the control set to determine whether to continue or exit the model iteration loop
4. **Classify:** Apply the completed model to the untagged project document population to classify each document as responsive or not responsive

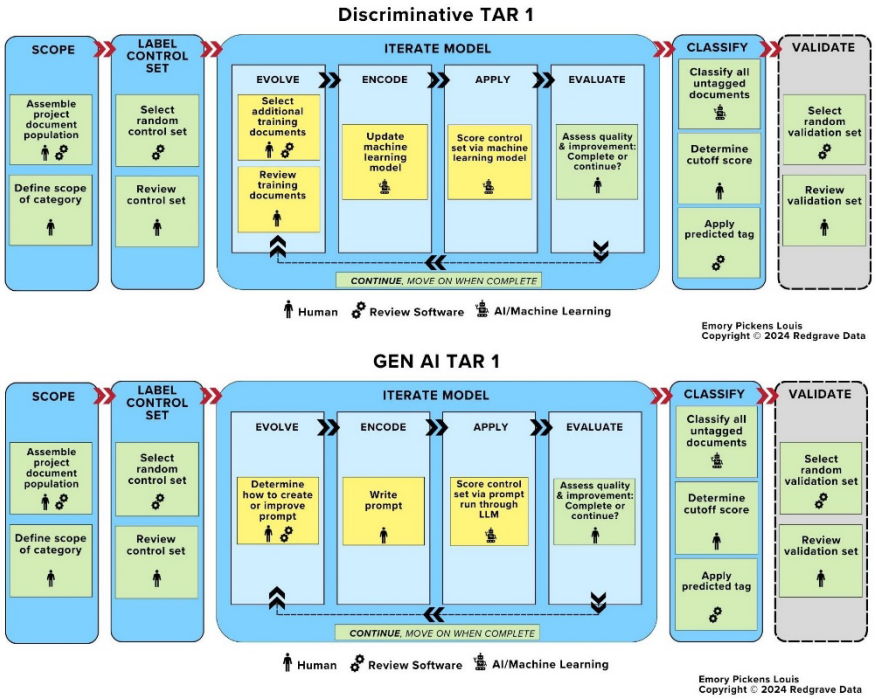
5. **Validate (optional):** Additional testing of the classified documents further evaluates the result

These steps reflect the TAR 1 process standard in eDiscovery. Its structure has enabled practitioners to efficiently and successfully use machine learning, expert systems, and now GenAI, to tag documents in a first-pass review, with metrics capable of demonstrating that the results meet requirements of reasonableness and proportionality. The conceptual steps of this TAR 1 reference model are not literal descriptions of every possible variation, and practitioners occasionally introduce slight modifications, without departing from core concepts.[9] At its essence, the steps of the TAR 1 Reference Model determine whether iterations are productive and improve the model, and whether the model's predictions are reasonably effective.

9. For example, one practice is to move the **Label Control Set** step into the **Iterate Model** step. In this minor variation, after each **Iterate Model** round, a new set of control documents are selected and reviewed, while the prior round's documents are used in the **Evolve** and **Encode** steps. For traditional discriminative TAR 1, this approach was initially used in some workflows. However, random samples (large enough to create sufficient certainty of model improvement) could be more efficiently taken and reused as controls sets, while active learning rather than random sampling during Model Iteration became preferred as a more efficient training approach.

D. WORKFLOWS FOR DISCRIMINATIVE TAR 1 AND GENAI TAR 1, COMPARED

The TAR 1 Reference Model illustrates a general TAR 1 workflow, which can involve different underlying algorithms. From this general model we can derive specific workflow diagrams for the tasks entailed when using traditional discriminative prediction engines versus when using GenAI.



Images reprinted with permission from Redgrave Data. Click on graphics for expanded view.

Through the reference model, the above workflow diagrams compare and contrast the tasks for workflows using traditional discriminative algorithms in discriminative TAR 1, and GenAI in GenAI TAR 1. The workflows are nearly identical; matching task boxes in the diagram are green to demonstrate consistency. They differ only in the **Iterate Model** steps of **Evolve**, **Encode**, and **Apply**, which we show with yellow task boxes, though the different approaches still accomplish the same five steps of the Reference Model. For convenience, we discuss these steps

below in context of a responsiveness review, although TAR 1 can be used for other purposes.

Step 1: Scope:

The project document set is assembled, and the scope of responsiveness is established. This step ensures that any subsequent human tagging or work done on the control set or the predictive model reflects substantive and statistical requirements to guide model iteration and prediction quality. If the responsiveness scope or project population changes after the project has begun, both the predictions and measurements may become incorrect and misleading.

Assemble project document population: The document set for the TAR review project should be selected.

Define scope of category: Attorneys determine the scope of responsiveness, defining each relevant issue or topic to be incorporated as responsive. The scope may be shaped by procedural requirements, facts known about the case, and requests for production. It is different from the document review protocol, which is based on the defined scope of responsiveness. While document review protocols may be adjusted for clarity and effectiveness, the scope of responsiveness should remain the same throughout a TAR project.

Step 2: Label Control Set:

The control set in a TAR 1 review functions as a test to estimate the predictive model's performance.[10]

Select random control set: A random population sample is selected from the TAR project population as a control set and remains independent from the training process. This sample provides an unbiased estimate of model effectiveness during the **Iterate Model** step. The larger the sample, the higher the confidence will be that the model's result for the control set is similar to its result on the entire review population. To ensure a model will be built that produces effective predictions on the general review population, the documents in that control set and any information directly gleaned (either by machine or by human) from those documents must not influence the Evolve step during **Iterate Model**.

Review control set: The control set undergoes human review. The tags of responsive or not responsive applied by the human reviewer will be compared to the model's predictions.

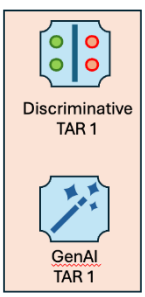
10. Control sets are a tool for the reviewing party to determine when proportionality considerations can limit their need to continue iterating the model to achieve better results. Control sets may not always be necessary, if a final validation demonstrates that the model was so effective as to leave little room for improvement through further iteration and make such efforts disproportional to potential benefits. Nevertheless, the purpose of the control set is to dramatically increase the probability that the final validation will be a success, so there is a chicken-egg consideration at play.

Step 3: Iterate Model:

A process loop is used to build a model that will predict the responsiveness of untagged documents. The **Iterate Model** step involves four sub-steps: a) **Evolve**, b) **Encode**, c) **Apply**, and d) **Evaluate**.

Other than excluding control set documents from the **Evolve** step, the TAR 1 training process is unrestricted on how and why documents and other information are selected, though practitioners should be mindful that some techniques will be more effective than others. Once the actual model building takes place, the model is then tested by applying it to the control set documents to estimate precision and recall of the results, and improvement (or lack thereof) compared to prior iterations. Then, a decision can be made about whether to continue iterating.

a. **Evolve:** This step involves improving the selection of information used to update a model's predictions.



Select additional training documents: Training documents are added as examples. Modern approaches to discriminative TAR 1 usually focus on selecting documents that will increase diversity (representativeness) of the examples or decrease uncertainty in areas where the model is most unsure. Examples can be selected in other ways, such as through random sampling or by human determination.

Review training documents: The new training documents are then reviewed and tagged by humans. This process may require a decent volume of training document review.

Determine how to create or improve prompt: A natural language prompt must be developed for the LLM. For an initial prompt, the prompt writer will need to consider the scope of responsiveness and information known about the document population to design the writing of an effective prompt. The writer needs to plan on how to instruct the system to analyze documents and determine responsiveness. The review software will typically specify the format for the LLM outputs (and automatically include those as part of the instructions to the LLM in the Apply step). For subsequent iterations, sources of information for potential improvements must be considered.[11] This

11. The work division between an LLM and humans in GenAI TAR 1 closely tracks that for expert systems, in which humans write "If-Then" rules for the machine. In place of If-Then rules, however, humans write natural-language, instructional prompts for the LLM. Both involve training the human, so the human can learn to write rules or a prompt to improve the model. For GenAI TAR 1 (and what we might call Expert Systems TAR 1), this occurs in the **Evolve** step. In contrast, discriminative TAR 1 trains the machine rather than the human, in the **Encode** step.

selection process is performed by a human, optionally assisted by machine-learning and information-retrieval algorithms. Example sources for prompt development can include custodial interviews and review of project documents (except those in the control set). The prompt writer should not be the same person who reviewed control documents or be exposed to information from the control set's contents, to avoid contaminating the objectivity and correctness of the control set.

b. **Encode:** The evolved information is incorporated into the model.

Update machine learning model: Using all training documents reviewed, the machine-learning algorithm updates the model, i.e. the machine learns.

Write prompt: Using the new information about how to improve a prompt that the human has now learned, the human rewrites the prompt.

c. **Apply:** The updated model makes predictions on the control set.

Score control set via machine learning model: A machine-learning algorithm uses the updated model to score the control set documents.

Score control set via prompt run through LLM: An LLM uses the prompt to generate a response for each control set document. The LLM's generative responses are then converted to scores or classifications.[12]

2. **Evaluate**: The model's performance is evaluated and a determination is made to complete or continue the Iterate Model loop.

Assess quality and improvement, and complete or continue: The control set scores are used to measure the quality of the model's predictions, typically with metrics of recall and precision.[13]

---

12. While not reflected in the diagram, review software will generally also submit its own instructions to the LLM to accompany the human's prompt, directing the LLM to provide responses in a format that the software can then map to classify each document. This will also be the case when the prompt is submitted in the **Classify** step.

13. Within a TAR workflow, recall is the percentage of all responsive documents found, out of all responsive documents in the project population. Precision in this context is the percentage of documents classified as responsive by the process that are actually responsive. *See The Sedona Conference Glossary: eDiscovery & Digital Information Management, Fifth Edition*, 21 SEDONA CONF. J. 263, 360–61 (2020) (citing The Sedona Conference, *Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 15 SEDONA CONF. J. 217 (2014) ("When describing search results, recall is the number of documents retrieved from a search divided by all of the responsive documents in a collection. For example, in a search for documents relevant to a document request, it is the percentage of documents returned compared against all documents that should have been returned and exist in the data set"); *Id*. at 354, ("When describing search results, precision is the number of true positives retrieved from a search divided by the total number of results returned. For example, in a search for documents

That quality is also compared to prior iterations to measure the extent of the model's improvement. Humans decide whether to continue the model iteration loop or complete it, based on whether the results are satisfactory, and the burden of additional model iteration loops is likely to outweigh benefits of continuing this process.

Step 4: Classify:

The built model is applied to the "real world" of the general project document population to predict responsiveness.

Classify all untagged documents: The completed model is run on all untagged documents to predict responsiveness. While this step is the same for both algorithms, the execution will be a little different. Discriminative TAR 1 algorithms' model will calculate a score (probability of responsiveness) for each document, while GenAI TAR 1 involves receiving a generated string of text from the prompt-fed LLM, which is then mapped to classification or to gradated scores. In traditional TAR 1 workflows using discriminative models, this step is commonly, but not necessarily, done at the same time as **Apply** in **Iterate Model.**

Determine cutoff score: For systems that predict with a score, whether discriminatively or

relevant to a document request, it is the percentage of documents returned that are actually relevant to the request").

generatively, a human determines the cutoff point. Based on the metrics of recall and precision from the control set, certain scores or predicted categories of documents can reasonably and defensibly be tagged responsive, and others not responsive. The higher the recall of the selected cutoff point, the more documents will be included, and the lower precision will be. When TAR 1 is used for production purposes, the cutoff point may be determined based on a legal requirement to meet a certain recall level (such as through a stipulated TAR protocol) or proportionality considerations of the value and burden of using different cutoff points.

Apply predicted label to documents: Based on the determined cutoff or predicted classification, the TAR software labels documents as responsive or not responsive.

Step 5: Validate (optional):

Additional sampling tests, through precision and recall, whether the model's extrapolation to the project population is as effective as was expected based on the control set. It validates the result rather than the model, which can eliminate subtle biases that may be introduced by repeated control set evaluation. However, this has historically not been standard practice in eDiscovery.

Select random validation set: A random population sample is selected from the document set that was classified by the model.

Review validation set: The validation set undergoes human review. The tags of responsive or

not responsive applied by the human reviewer are compared to the tags applied by the model to assess the results of the TAR 1 process.

E. CONSIDERATIONS FOR GENAI VERSUS DISCRIMINATIVE ALGORITHMS IN TAR 1

GenAI TAR 1 introduces promising advantages that could make it an important tool for eDiscovery practitioners. Studies are needed in several areas to assist practitioners in evaluating whether and when to select GenAI TAR 1, discriminative TAR 1, other workflows, or even hybrid approaches that blend GenAI TAR 1 with other workflow options. This field is still in the early stages of evaluation. Initial studies have tested the effectiveness of GenAI predictions for document tagging in various ways, including against humans, against discriminative algorithms, and against other LLMs.[14] There are not yet studies comparing fully iterated TAR 1 workflows using GenAI versus using discriminative algorithms, though they will surely come in time. Specific issues for study, and that practitioners should consider to decide what will best serve their needs, should include 1) precision and recall, 2) risk of sensitive information

14. *See* ROSHANAK OMRANI, ET AL., BEYOND THE BAR: GENERATIVE AI AS A TRANSFORMATIVE COMPONENT IN LEGAL DOCUMENT REVIEW, Relativity and Redgrave Data (Feb. 2024) (comparing to manual review); Colleen M. Kenney, Matt S. Jackson & Robert D. Keeling, *Replacing Attorney Review? Sidley's Experimental Assessment of GPT-4's Performance in Document Review*, THE AMERICAN LAWYER, https://www.law.com/americanlawyer/2023/12/13/replacing-attorney-review-sidleys-experimental-assessment-of-gpt-4s-performance-in-document-review/?slreturn=20240204151012 (Dec. 13, 2023) (comparing to manual review); SUMIT PAI, ET AL., EXPLORATION OF OPEN LARGE LANGUAGE MODELS FOR EDISCOVERY, Proceedings of the Natural Legal Language Processing Workshop (Dec. 2023) (comparing different LLMs), *available at* https://aclanthology.org/2023.nllp-1.17.pdf; JASON R. BARON, NATHANIEL W. ROLLINGS & DOUGLAS W. OARD, USING CHATGPT FOR THE FOIA EXEMPTION 5 DELIBERATIVE PROCESS PRIVILEGE, Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (June 2023) (comparing discriminative and LLM performance for FOIA deliberative process privilege analysis), *available at* https://ceur-ws.org/Vol-3423/paper4.pdf.

disclosure, 3) knowledge gain and accomplishment of related tasks, 4) total project cost, 5) total project time, 6) ease of use, and 7) whether different algorithms may best serve different needs.

1. **Recall and precision:** Precision concerns may be somewhat alleviated if either GenAI TAR 1 or discriminative TAR 1 is found to be more precise than the other at similar defensible recall levels.[15] Use of TAR 1 has been limited in part because when human review is skipped and documents that are predicted responsive are then produced, significant numbers of not-responsive documents are often included. This limitation may resolve if GenAI can reduce that risk by achieving higher precision (fewer nonresponsive documents in the production set) at the same or higher recall rates (finding as many or more responsive documents) compared to discriminative TAR 1.[16] On the other hand, unlike discriminative TAR 1, GenAI TAR 1 may only classify at a few gradations of responsiveness, which may require selection of cutoff points with lower-than-desired responsiveness in some cases.[17]

---

15. To the extent even higher recall may be achieved with GenAI with high precision, burdens of producing at higher recall levels may be reduced and create a win-win in which producing parties may face less risk while producing even more responsive documents to receiving parties, as compared to current industry practices.

16. Metrics are essential to determining effectiveness of any process, including GenAI. Although some practitioners may be more accepting of GenAI TAR 1 than discriminative TAR 1 because GenAI can "explain" its decisions, those decisions are also predictive text and do not inherently give credibility to its classification predictions.

17. Discriminative algorithms produce real-valued scores that allow very fine gradations that often can near-uniquely rank all documents in the project population, from those predicted most to least responsive. This facilitates

2. **Sensitive information:** Another hesitation for traditional TAR 1 workflows has been the risk of sensitive information disclosure, such as privileged or personal information, when it is used to produce documents without human review. While this risk is mostly managed with additional workflows, such as keyword screens for privileged information with human review, it is always possible that important information was missed. But, in addition to predicting documents as responsive or not during a TAR 1 workflow, GenAI may simultaneously also be able to identify confidential and sensitive information, saving costs and time, if it can do so with similar effectiveness as traditional approaches.
3. **Knowledge gain and multitasking:** Because TAR 1 applies tags without human review, it is known to do little for case teams in terms of gaining knowledge and insights about the documents from first-pass reviewers. As discussed, this limitation of TAR 1 may be lessened or overcome, because GenAI can also create summaries and identify key points in documents as it also predicts responsiveness, which may then be useful for case teams. It may be able to simultaneously accomplish other tasks

a practitioner's selection of a cutoff score based on many options with different recall and precision scores (e.g., produce at 75% responsiveness with 55% precision, at 80% responsiveness with 45% precision, or many other options in between). In contrast, if GenAI TAR 1 only classifies documents into a few gradations of responsiveness (e.g., binary as responsive or not, or on a small scale), practitioners may have to choose between a cutoff point with very high recall and low precision, or low recall and higher precision, with no choices in between. For example, given a result with few gradations, assume two result points, with one of 50% recall and 90% precision, and the next possible cutoff point option with 98% recall at 25% precision (which may contain the vast majority of the review population). With no options in between, one may be faced with only two unhelpful choices.

discriminative TAR 1 does not, further saving time and costs over those workflows, such as identifying privileged material and personal information, identifying and reviewing foreign language documents, and privilege logging. It remains to be seen whether such additional uses of GenAI in conjunction with predictive uses can be deployed in ways that make it more useful and cost effective than human first-pass review.

4. **Costs:** Total costs of GenAI TAR 1, including direct costs of the tool as well as attorney (prompt iteration) and litigation support costs, need to be evaluated against other options. Currently, compared to discriminative algorithm use, each document reviewed by GenAI will be at a considerably higher cost. In many cases, discriminative TAR models can be iterated and applied without limit and without incurring additional costs. In contrast, GenAI is more expensive, and may involve additional costs for every prompt sent and answered in the steps of **Iterate Model** and **Classify**, at least one for every document in the project population. However, GenAI TAR 1 also has the potential to save on costs from attorneys and support staff. Discriminative TAR 1 generally requires training with several thousand documents that are often, though not necessarily, reviewed by subject-matter expert (high-cost) reviewers. This effort should be compared to the cost of any document review and other efforts that will be required for effective prompt writing and iteration in GenAI TAR 1 .
5. **Time:** The total speed of project completion should be considered, including both human and machine time. Workflows must fit circumstances of case needs and deadlines, so timing matters. Currently, machine time for GenAI TAR 1 is slower than discriminative TAR. However, this will likely improve and may also be offset if the

prompt development process is faster than reviewing training documents for discriminative TAR training, as discussed above. Though again, some review of documents will likely be required during GenAI TAR 1 prompt iteration as well.

6. **Ease of use:** GenAI TAR 1 may be preferred by practitioners if it is easier and more practical to use. To begin with, the process of writing a query may feel more approachable, and not require much instruction to try, as opposed to most discriminative TAR systems. In addition, attorneys generally do not relish reviewing thousands of documents to train discriminative TAR 1, and GenAI may save them this task if prompt writing and development is easier. On the other hand, reviewing documents is not a challenging task, and the comparative ease of successful prompt writing and iteration is still unknown.[18] Some may be dissuaded from GenAI TAR 1 adoption if it presents less certainty of success, costs, and time; this may also be affected by the skill of the prompt writer. In addition, GenAI TAR 1 requires the prompt writer to be a different person than the control set reviewer, and shielded from exposure to the control set, which may limit its practicality for some case teams.
7. **Other considerations:** Practitioners should keep in mind that even studies on the above issues do not guarantee their own project will always have similar results. It may be that different circumstances affect outcomes, such as type of matter, document volumes, nature of responsiveness and issues, case team composition, and other factors.

18. In some circumstances, it may be easier to recognize whether a document is responsive than it is to describe all aspects of responsiveness.

F. HYBRID WORKFLOWS: MIXING ALGORITHMS

GenAI may be integrated with other eDiscovery tools to yield even more possibilities for TAR 1 and other workflow improvements. Hybrid approaches are already common in eDiscovery workflows generally, with mixtures of machine learning, search terms, conceptual search, and structured (metadata) analytics. The most effective use of GenAI, including but not limited to its application in GenAI TAR 1, may involve integration with other approaches.

In fact, GenAI as used for query responses is already a hybrid of processes, as it often leverages a type of combination workflow known as RAG (Retrieval Augmented Generation). This approach involves a (nongenerative) conceptual search of the query against the document set to find the most closely related documents. Those documents are then fed to the LLM along with the query as a prompt, and the LLM then produces a response based on those documents.

The process of GenAI TAR 1 may similarly benefit from other traditional systems. The process of evolving a prompt may be enhanced by discriminative algorithms, as well as diversity algorithms, which may identify documents that will be most helpful for the prompt writer by expanding their knowledge of "unknown unknowns" in the project population. Additionally, where GenAI TAR 1 predictions may designate large volumes of documents as responsive but only provide limited gradations of responsiveness, traditional discriminative algorithms based on a modicum of training may fill in the gaps by providing within-gradation secondary scores.

Conversely, GenAI used more broadly may be able to generate content useful to assist TAR 1 training for both discriminative and GenAI TAR 1. For example, it could generate search queries to retrieve potentially useful training documents, create synthetic training documents, or tag documents to train a

discriminative algorithm. When using GenAI TAR 1, it can help prompt writers evolve their prompts by asking questions to identify and clarify unintended ambiguities in the defined scope of responsiveness.

With GenAI as another tool in the belt of eDiscovery practitioners, new and creative applications will continue to appear. However, novelty must be accompanied by evaluation if it is to become innovation; just because something can be done does not mean it will produce a better outcome than a related, known-effective approach, no matter how plausible the novel idea seems.

G. CONCLUSION

As illustrated in the TAR 1 Reference Model, established and defensible processes for predictively tagging documents through a TAR 1 process follow steps of: 1) **Scope**, 2) **Label Control Set**, 3) **Iterate Model**, 4) **Classify**, and 5) (optionally) **Validate**. While GenAI, in the form of LLMs, offers new possibilities for improving the efficiency and effectiveness of first-pass document review, its use still follows the established steps of TAR 1. This process, which involves sampling and statistics, will help promote successful outcomes on first-pass review projects for practitioners using GenAI — as it has helped those same practitioners when using discriminative approaches.

Especially as GenAI capabilities increase and costs and time it requires go down, GenAI has potential to become a preferred approach to TAR 1, as discussed above. Future studies may demonstrate that GenAI can be a more effective choice. To improve TAR 1 workflows, and relative to discriminative models and not just to linear review, GenAI will need to achieve improved recall and precision, effectively and efficiently incorporate other tasks that go beyond first-pass review, be cost effective and sufficiently fast, and be practical for case teams to use. In addition, the potential to mix GenAI with other algorithms into hybrid approaches may further increase its value in improving first-pass review.

Guided by the structured approach of the TAR 1 Reference Model, practitioners have much to consider in selecting approaches, given the well-known benefits and risks of discriminative TAR 1, the untested but potential capabilities of GenAI TAR 1, and the option to mix algorithms. Approaches to document review may change significantly in some ways with the incorporation of GenAI, but they will also be fundamentally unchanged in others.