DISCO Prioritized Review
Best Practices

With the assistance of powerful technology such as DISCO AI, maximize the work of attorneys and review teams to provide your client with the most cost-effective review.

DISCO

Estimation Sample → **Targeted Review** → **Document Review** ⇄ **CAL** → **Validation Sample**

Understanding DISCO AI[1]

DISCO artificial intelligence (DISCO AI) is a next-generation continuous asynchronous learning (CAL) technology designed to amplify the work of attorneys and review teams. It can be applied to your cases to:

- Reduce the volume of review required to complete a document production

- Prioritize document review to help reviewers identify the most relevant content quickly

- Improve the effectiveness of fact identification for document investigations

- Identify overlooked documents that may be valuable for functions such as deposition prep

- Assist with quality control of privilege review, reviewer consistency, and general review accuracy

This document provides an overview of best practices for DISCO AI to increase the efficiency of a production preparation or examination. While DISCO AI is a highly flexible and easily accessible tool, we nevertheless encourage you to contact your project manager for a consultation on the best ways to tailor this tool to your specific case goals.

To understand DISCO AI, it is often easiest to relate it to similar consumer technologies such as popular audio or video streaming services. These products are designed to learn from your likes and dislikes with the simple goal of recommending content that you may enjoy. Often this is facilitated by an easy thumbs-up-thumbs-down ranking system. The more you consume and interact with the media, the better the service learns your interests and makes smarter recommendations. DISCO AI is very similar with respect to understanding your tagging behavior. As you tag documents for categories such as *Responsiveness, Issues,* or *Privilege,* DISCO AI will evaluate the content of all your documents and help organize the entire population based on which records are likely or unlikely to be related to the existing tagging decisions.

---

[1] While this paper is geared toward incorporating DISCO's tag predictions in your case review process, the majority of steps laid out within this paper can and should be followed for all reviews within DISCO.

DISCO identifies likely or unlikely content by applying a predictive tag score to each document. Once a DISCO AI-enabled tag has been applied to 50 documents, DISCO starts to score all the documents for that tag. Tag scores range from -100 to +100, depending on how likely or unlikely DISCO believes a certain tag is predicted to be applicable to a certain document. When used in conjunction with review stages, DISCO predictive scores enable clients to easily and efficiently conduct a prioritized review. By allowing you to prioritize your review with tag prediction score(s), DISCO intelligently front loads the review with the most relevant content.

DISCO AI's scoring model relies on positive and negative signals. A positive signal is sent when a tag is applied to one or more documents (including during bulk tagging). Negative signals are sent when DISCO detects the choice to not apply a tag. Since CAL powers DISCO AI's predictions, DISCO AI learns from each tag you apply — as well as each tag not applied — throughout your review process. Furthermore, DISCO continuously recalculates scores, ensuring that you are always working with the most up-to-date model. When used in conjunction with review stages, the continuous recalculations ensure that you are continually reviewing documents that are most likely to meet your criteria.

A negative signal is captured only in instances when another tag is applied. (Example: A reviewer has Issue A, Issue B, and Issue C on their tagging pane. The reviewer tags only Issue A. This sends a positive signal for Issue A and negative signals for Issue B and Issue C.) Unlike positive signals, negative signals are not captured in bulk-tagging operations. Because DISCO AI works most optimally when receiving both positive and negative signals, we recommend configuring your review — and your tagging pane — with this dynamic in mind.

Preparing your data population for a DISCO AI review

As with most reviews, it is essential to identify which content is relevant to your objectives. Often this process is referred to as early data assessment (EDA). Parties may jointly or unilaterally decide to cull content based on functions such as date restrictions, search term results, and file type considerations. It is best practice to decide upon these parameters before commencing with any review. DISCO's search term reporting (STR), search filters, and search visualization features all provide useful insight to help inform such decisions. DISCO folders serve as an excellent repository for storing your review population along with any other categories of documents uncovered during your exploration. For more detail on EDA techniques within DISCO, please contact your project manager.

Setting up predictive tags for your matter review

DISCO is ready for review with minimal set-up. The platform provides various functions to support review efforts such as term highlighting, foldering, and review workflows. For the purposes of this white paper, we will focus on tags, the key driver of DISCO AI.

To get you started, DISCO pre-populates each matter with the most commonly used tag groups such as responsiveness, issues, and privilege. You can easily add or remove these tag groups or specific tags to meet your case needs. When creating tags, it is essential to consider what information you want your reviewers to capture. For example, when reviewing an opposing counsel's production, you might want to replace responsive and non-responsive tags with relevant and non-relevant tags. Additionally, you may want to add confidentiality tags to designate various levels of agreed-upon protection, along with issue-specific tags to capture key information that might be used later in preparing for depositions or trial. It is recommended to always have reviewers apply at least one tag during the review — this confirms that the document has been examined.

All Tags

NAME	GROUP	DOCS	SHORTCUT	PREDICTIONS
Add new tag...	Select or create group ▼	+		
Enron - Oil, Gas or Energy	Enron Predictive Tags	5648	SHIFT + 1	ON
Enron Outside USA	Enron Predictive Tags	3913	SHIFT + 2	ON
FERC	Issue	11784	SHIFT + F	OFF
Further Review	Responsiveness	19	SHIFT + U	OFF
Hot	Importance	101	SHIFT + H	OFF
Inappropriate HR Content	Enron Predictive Tags	259	SHIFT + 5	ON
interesting	Responsiveness	1	SHIFT + S	OFF

Setting up predictive tags for your matter in DISCO is easy. Simply toggle on the predictions button and DISCO begins monitoring the tag for both positive and negative tagging behaviors.

Consider which tag(s) you will use for your predictive prioritized review when creating tags. While DISCO AI can be turned on at any time during a case, it is beneficial to feed the algorithm from the beginning of the review. DISCO AI learns continuously, so any changes in tagging behavior (resulting from, for example, a change in review protocol or improvement of reviewer performance) will quickly be absorbed into the intelligence of tagging predictions. Generally speaking, it is best to use discrete issue codes and turn predictions on for any issues that you want to score and prioritize by. To turn DISCO AI on, flip the predictions toggle to on in management.

It is important to consider the application of tag(s) and we encourage document-level coding during your review. Meaning, you should apply tags to a document that reflect the content of the document rather than the content of any related documents, such as family and conversation members.

While we encourage reviewing within the four corners of each document, you can still leverage the advantages of tag propagation in Review Stages. For example, if you choose to set your family inclusive review stage propagation to "Document and attachments." When a reviewer tags a parent document as Responsive and marks it as reviewed, a positive signal for Responsive is sent for the parent document. When the reviewer navigates to the attachment, they will see the propagated responsive tag. However, a signal for the Responsive tag on the attachment, will not be sent to the AI until the reviewer has marked the attachment as Reviewed. At this point, because the reviewer took an explicit action, the propagated tag will inform machine learning.

Estimation sampling

Unlike many other platforms, using DISCO AI does not require you to first train DISCO. However, we do recommend conducting a estimation sampling to determine prevalence as a prudent first step in any review, and especially when you intend to leverage DISCO's predictive analytics. Although it is not required, it is often recommended that subject matter experts review all important samples as this may help reinforce the defensibility of your review should you choose not to review the entire corpus. Samples are a helpful tool for several reasons:

- Taking a sample allows you to forecast the prevalence of responsive documents across your review population, along with privilege and issue tagging.

- A sample can give you a representative preview of your data set, giving you insight into key players and issues in the review.

- A sample can provide early insight into the over- and under-inclusiveness of proposed search terms.

- Reviewing a sample starts to send signals to DISCO's machine learning system, allowing it to begin building predictions even before the full review begins.

To start, you need to select the degree of confidence and margin of error you want to get from your results. We generally suggest aiming for a 95% degree of confidence with a 2% margin of error. To calculate the precise number of documents you will need to review, you can use any statistical calculator, including the one built into DISCO's quality control (QC) feature. For example, https://www.surveymonkey.com/mp/sample-size-calculator provides an simple online calculator to determine your necessary sample size.

DISCO provides you with the following search syntax that will quickly and easily gather a random sample of documents:

```
sample({size}, {population})
```

For example, if we had placed the documents slated for review into a folder called *Potentially Responsive,* the syntax to pull our statistical sample would be as follows:

```
sample({size}, folder("Potentially Responsive"))
```

If, on the other hand, we wanted to pull the sample from the entire database, our syntax would be:

```
sample({size}, *)
```

Each time you run the sample search syntax, it will yield a different set of documents. Therefore we recommend placing your search results into a folder, which will allow you to return to your exact sample set in the future.

Once the review of your sample set is complete, you can make forecasts about your document population. In the above scenario, we selected a 95% confidence level, with a 2% margin of error. This means that if, during your sample review, you tagged 17% of the documents as responsive, you could expect, with 95% certainty, to find approximately 15%–19% of your underlying review set to be responsive. These estimates can be used to help inform your decisions regarding the timing and the number of reviewers needed to review your documents. Additionally, by examining the differences between what you expected to find and what you actually found, you may gain useful insight regarding your matter's evolution.

Note: The results of a statistical sample will not be dependable with any significant change to the review corpus (adding or removing documents). DISCO AI is highly flexible to this scenario; however, it is typically recommended to review a new sample representative of the full corpus whenever significant changes are made.

Targeted review

In any review, the question is always, "Where do we start?". You may have already identified important content during your EDA process. Since DISCO AI is designed to learn from you and amplify your efficiency, it is best to start the review based on targeted sets of data known to likely be responsive to the matter. This can be something as simple as starting with a priority custodian, search terms and/or date ranges. Just as you did with cases before using DISCO AI, apply what you know about the case and run searches to find some starting points. To note, a targeted review should occur after the estimation sample is complete as sampling a targeted population will not be indicative of the entire corpus.

Keep in mind powerful targeting functionalities such as STR, search filters, and search visualization. Similar to using these for the aforementioned EDA process, when creating targeted review data, these tools create great focus and visibility as to where you begin. Saved searches and document folders help to easily retrieve documents and allow you to easily incorporate them into a review stage. With DISCO's Review Stages, it is possible to integrate your statistical sample and targeted searches into a prioritized review.

DISCO review stages should be used for any review where you intend to do one or more of the following:

- Make the fullest use of DISCO AI
- Track your review and forecast timing of review completion
- Use a condition of *review completed or reviewed* by *(reviewer)* as search criteria
- Divide the review population into batches
- Organize levels of review (first pass, second pass, privilege review, etc.)

The configuration of your review stages will determine not only which documents are reviewed, but how they are ordered within batches and what tagging decisions can be made.

Defining your sources

Sources determine which documents are included in your review stage. Use the tags, folders, and saved searches created during the exploration of your data to define your review set accurately. You can also create ad hoc searches as you setup your source. You may have up to 20 discrete sources per review stage, each containing a distinct search string.

Review stages are dynamic by design. DISCO uses source criteria, such as searches or predictive scores, and continuously runs that criteria across the entire document population. While this is very helpful for ensuring that the highest scoring tag predictions consistently move to the front of your review, it might not be desired in every case. Therefore, if you want to limit your review to a specific set of documents (for example, disregarding later data ingests), you will have to restrict your source by utilizing a folder or tag to create a static population to feed into the review stage.

Once you have created the sources for your review stage, you will decide whether or not to include families and/or conversations. Check the appropriate box(es) to ensure that all family members and/or members of a conversation are included in the same batch. Including the entire conversation enables DISCO to gather all documents from a conversation thread within a single batch for an attorney to logically and consistently review. It should be noted, however, that including conversations may expand the review universe to include documents that were not originally in the source searches.

Next, choose how you want to group the documents within your review stage. These options dictate how the batches will be created and prioritized. For example, you can prioritize certain custodians, ingest sessions, or sources to be reviewed first.

Defining your sources in DISCO determines which documents are included in your review stage. You may have up to 20 discrete sources per review stage, each containing a distinct search string.

Prioritizing review based on DISCO AI

There are multiple approaches when performing a review based on a CAL model. In every scenario, DISCO AI will prioritize (or front load) your review batches with content most likely to be responsive.

A simple use case is to prioritize based on the *Responsive* tag itself. When looking to create a workflow where the documents will be prioritized strictly by a tag's predictive score, the No groups option should be selected, and you will order by tag predictions instead. Ordering of the review stage determines the logical order that the documents will appear within batches. For instance, if you want to review the documents that DISCO has identified as most likely to be tagged *Responsive* first, set the order by to tag prediction: *Responsive* tag in descending order. This will push all documents with high *Responsive* tag predictive scores to the front of the review.

To gain more efficiencies, DISCO project managers often recommend creating a prioritized workflow based on multiple issue tags related to responsiveness. Since documents of the same issue are more likely more to be similar to one another, the machine learning may return better recommendations more quickly.

Regardless of approach, we recommend turning predictions on for all key work-product tags so that you have the flexibility to adjust review strategy quickly.

Creating your decision pane

Once you have chosen the sources and sorting for your review stage, it is time to create the decision pane. For redactions, document notes, privilege notes, and any custom fields, add them to the decision pane. For tags, choose to either add all tags (as seen in the standard document viewer) or create logical tag groupings with either single-select or multi-select criteria. Keep in mind that any tags included in the decision pane should be part of your evaluation for each document.

For optimal use with DISCO AI, we recommend using pre-set tag groupings. For example, it is best practice to group together the Responsive, Non-responsive, and Further Review tags, and require reviewers to select at least one option. Then it is recommended to place Privilege tags and Issue tags within their own groups and allow for multiple selections from each group. Remember, an administrator can mark any item placed within the decision pane as required. If a decision is required, reviewers will not be able to mark the documents as reviewed until the required decision is made. As noted above, we recommend making at least one tag required, which allows for better identification of reviewed documents outside of review stages.

Just-in-time batching

DISCO uses unique just-in-time batching. Batches are generated at the moment a reviewer requests a new batch. This tremendously reduces administration time and makes it easy to keep reviewers working on your highest-priority data. For example, you can quickly re-prioritize the order of custodians in your review stage without having to remove and re-generate a set of batches. New batches will reflect your updated criteria without causing any administrative disruptions.

Just-in-time batching is particularly helpful when conducting a predictive prioritized review. If you prioritize tag recommendations for a certain issue, DISCO will steadily push the highest-recommended documents to the front of your review. Furthermore, as DISCO AI's predictions become stronger, each new batch checked out will contain the documents with the highest available DISCO AI scores from the remaining unreviewed documents.
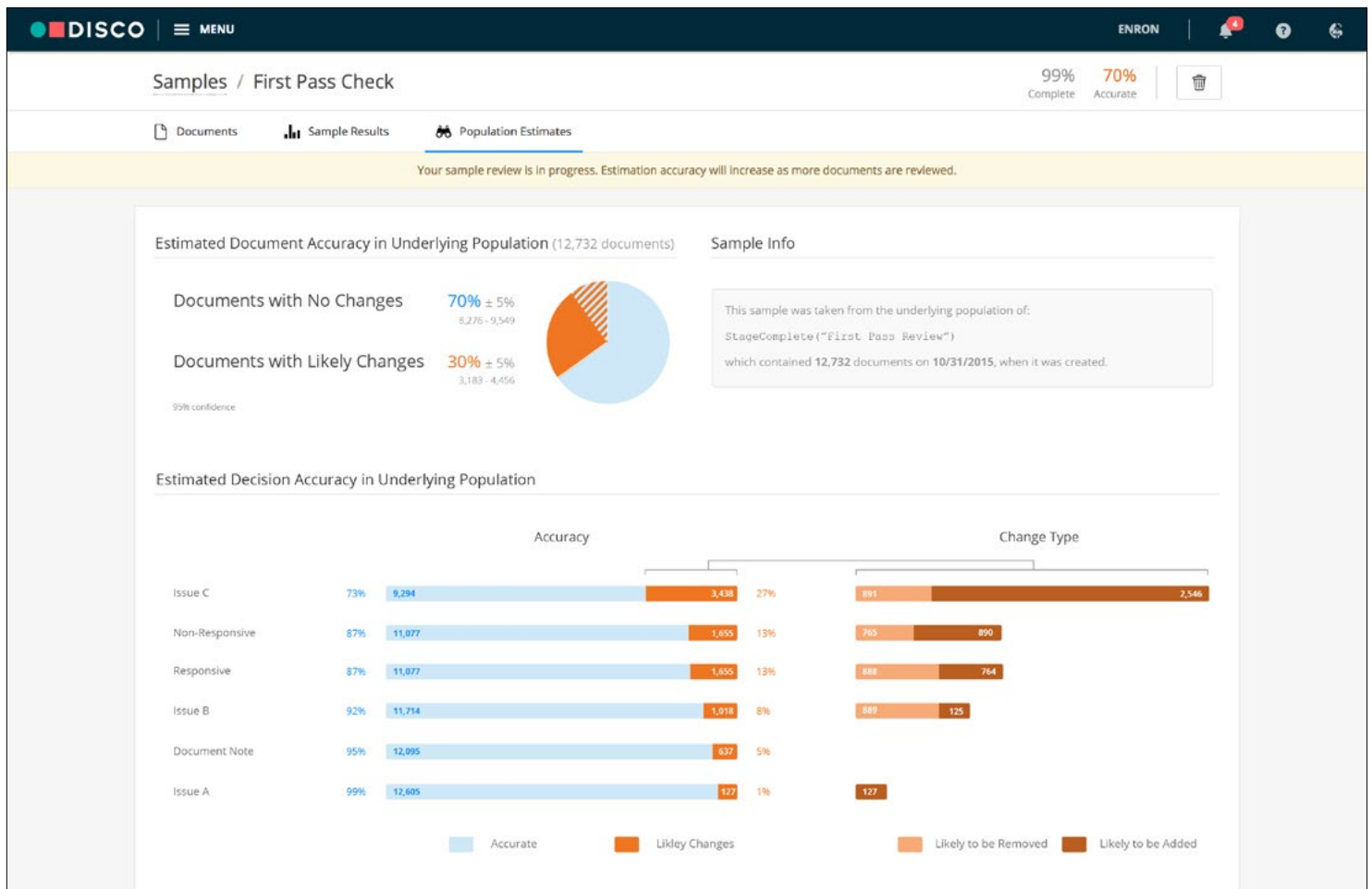
Tracking your review

The review stage metrics tool allows you to track the pace of each review stage and to analyze and forecast review progress. This is particularly important when utilizing a predictive prioritized review. Because DISCO is continuously learning from your tagging behavior, you want to feed it the most accurate information possible.

- **Pace:** Displays the team's overall review pace along with the median pace of active reviewers. DISCO will estimate how long the review will take based on the current review pace.
- **People:** DISCO provides charts that show reviewer pace by day and tagging rate by reviewer. These charts provide insight into the overall review, and highlight outliers among your review team.
- **Findings:** View the percentage distribution of tags applied within the stage. Double click on any specific tag to review the associated documents.

Quality control

As you move through the review, use the information provided in the metrics feature to determine if you need to conduct an accuracy check. DISCO QC allows you to spot check the accuracy of the entire review or individual reviewers, and statistically infer the accuracy of the reviewed population as a whole. Simply select your QC set based on criteria such as stage name, tag name, date of tags application, or even reviewer name. Next choose which review decisions (tags, redactions, or notes) will be evaluated and recorded as overturns. Then select your confidence and margin of error levels.

Based on the results of the quality control sample, DISCO will project a detailed breakdown of the accuracy of the sampled population, including estimates of how many documents will likely require coding overturns, and the type of overturns that will likely be required.

Once you have chosen your criteria, DISCO will gather a random sample of documents for review. During the QC review, DISCO will track each time you add/remove a tag, note, or redaction and track each change as an overturn. As you move through the QC document set, DISCO will display the percentage complete for your QC review, as well as the accuracy percentage ascertained through your QC review to date.

Based on the results of the QC sample, DISCO will project a detailed breakdown of the accuracy of the sampled population, including estimates of how many documents will likely require tagging overturns, and the type of overturns that will likely be called for. Periodic QC sampling can catch problems early on in the review and provide confidence in your team's tagging decisions.

DISCO AI's tag prediction scores can also be utilized for further quality assurance. For example, you can search for documents tagged *Responsive* by a reviewer, but for which the prediction score is strongly negative (and vice-versa). By performing quality checks on documents with a large disparity between their tag predictions and their actual tagging, potential problems related to reviewer quality, changes to responsiveness guidelines, and more can be identified.

Validation sampling

As your review progresses, you can use the information generated from the statistical sample along with DISCO AI's tag predictions to help determine when the review is complete. For example, if the number of responsive documents that have been found is close to the number forecasted by the prevalence within your sample, or if DISCO AI no longer recommends any additional documents (per the predictive scoring), consider running a validation sample of the remaining unreviewed documents.

For the validation sample, you may want a higher degree of confidence and a lower margin of error, since you may use this sample to defend your review. An acceptable number might be 99% confidence with a 2% margin of error. Again, you can use DISCO's sample search syntax and review stages to conduct your validation review. Your findings will help you determine whether you need to continue the review, whether you can defensibly stop the review, or whether you want to complete the review using a more cost-effective team (since it is likely that most of the responsive documents have already been located). It is also good to remember that, in addition to sampling the unreviewed set, you can still leverage DISCO's search and data analysis tools to try to target potentially responsive documents within the population.

Conclusion

For defensibility purposes, it is important to diligently and consistently document key decisions regarding your review workflow. Maintain records and lists of keywords, custodians, date ranges, and other considerations used for culling, sampling, and review. While DISCO's arsenal of tools can greatly assist lawyers in reviewing documents, **the decision whether to continue or stop the review should be based on the legal judgment of the attorney managing the review.** With a combination of powerful technology and a well-contemplated, well-documented strategy, you will be able to provide your client with an efficient and cost-effective review.

Additional Resources

Maximizing Attorney Intelligence: Solutions Towards a Just, Speedy, and Inexpensive Case Resolution by Trevor Jefferies

Practical Advice on How to Run a Predictive Prioritized Review by Trevor Jefferies

Summary sheet

DISCO Prioritized Review

Estimation Sample → **Targeted Review** → **Document Review** ⇄ **CAL** → **Validation Sample**

Prepare your data

- Use DISCO Search Term Reporting (STR), search filters, and search visualizations to identify the review population
- Folder or tag the likely relevant documents for review

Prepare your tags

- Create any tags for issues or privilege as needed
- Toggle on the 'Predictions' switch on the tag management page for all the tags for which you would like to leverage tag predictions

Estimation sampling

- Understand the prevalence of responsiveness, privilege, and issues across your review population
- Draw a random sample of documents from the review population that will give you 95% confidence with a 2% margin of error
- Place these documents in a folder and review them
- If there are significant changes to review population (e.g. new documents added), restart this step

Targeted review

- Review a targeted set of documents based on relevant searches, date ranges, custodians, keywords, etc. to help DISCO AI learn from your tagging behaviors
- Use STR, search filters, and search visualization to identify the set of documents most likely to be responsive to the tags for which you are reviewing
- Draw a random sample of documents from this subset
- Place them in a folder and review them

Document review

- Use DISCO review stages to review the remaining documents
- Define your sources
- Prioritize batches based on predictions from DISCO AI
- Customize your decision pane
- Track your review progress and accuracy
- Perform periodic quality control using DISCO's QC feature

Validation sampling

- Identify when you have reached the end of your review and when you're likely to not find many more relevant documents by continuing to review
- Draw a random sample of documents from the unreviewed document population that will give you a 99% confidence with a 2% margin of error
- Place these documents in a folder and review them
- Based on your findings, you can decide if you would like to continue reviewing the remaining documents