



Best Practices: Technology Assisted Review

Applying Transparent, Scalable Predictive Coding Technology to Speed Document Review and Reduce Costs

April 16, 2015

eDiscovery
WebinarSeries

Karsten Weber
Lexbe LC

eDiscovery Webinar Series

Info & Future

- **Takes Place Monthly**
- **Cover a Variety of Relevant eDiscovery Topics**
- **Presentations Available for Download by Registrants.**

eDiscovery Webinar Series

About Lexbe

Lexbe is an Austin, TX based eDiscovery software and services provider.

- **Lexbe eDiscovery Platform**

Lexbe eDiscovery Platform is a hosted eDiscovery processing and review tool. Users can load a variety of file types, process for review, OCR for search, and conduct document reviews, productions, prepare for depositions & analyze transcripts, conduct case analytics, prepare for dispositive motions, and provide litigation support during trial. Per GB hosting charges won't break the bank and there are no user fees.

- **Lexbe eDiscovery Services**

Lexbe does large volume document culling, processing from native to PDF or TIFF, load file creation, high-volume OCR of image files, Rule 26 and project management consulting, and related eDiscovery Services.



Lexbe is recognized as a 'Top 100' eDiscovery Provider by ComplexDiscovery, a leading electronic discovery and information governance firm.

**eDiscovery
Webinar Series**

Lexbe Sales
sales@lexbe.com
(800) 401-7809 x22

lexbeSM

eDiscovery Webinar Series

Questions & Technical Issues

If you have any questions or technical issues, please e-mail them to:

webinars@lexbe.com

eDiscovery
Webinar Series

eDiscovery Webinar Series

Karsten Weber bio

- **Current**
 - Principal of Lexbe LC
 - Principal Architect of Lexbe eDiscovery Platform and Lexbe eDiscovery Services
- **Prior Experience**
 - Consulting Expert, Lumin Expert Group
 - Director of Software, nLine Corporation
 - Software Engineering Manager, KLA-Tencor
- **Education**
 - MBA, University of Texas
 - M.S. Engineering, Danish Technical University

Contact

Karsten Weber
512-686-3469
karsten@lexbe.com

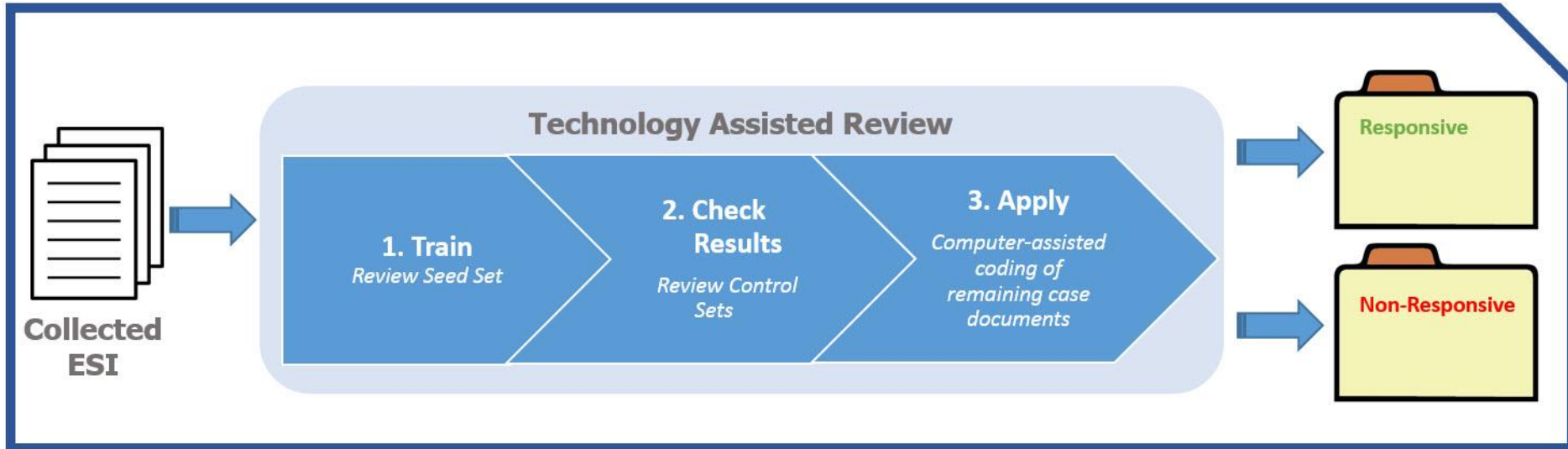
Best Practices: Technology Assisted Review

Agenda

- What is Technology Assisted Review (TAR)?
- How does TAR/Predictive Coding work?
- Why use TAR/Predictive Coding?
- Comparing outcomes: predictive coding vs. and manual review
- Importance of transparency in TAR applications
- Benefits of scalability in predictive coding architectures

Best Practices: Technology Assisted Review

What is TAR/Predictive Coding?



- Predictive coding allows a skilled reviewer to train a computer algorithm to identify responsive and non-responsive documents in a litigation document collection.
- As an alternative to manual linear review, predictive coding can drastically reduce the amount of time needed to review increasingly large ESI volumes.

Best Practices: Technology Assisted Review

Why Use TAR/Predictive Coding?

Increase Review Speed: TAR is designed to complete review of large ESI collections faster than human reviewers. Applying TAR in a scalable environment maximizes the speed advantage of predictive coding.

Decrease Review Costs: Whether paying per document or per hour, TAR is significantly less expensive than exhaustive manual review.

Increase Review Quality: Many studies conclude that the presumed quality advantage of 'gold-standard' manual review is not accurate. TAR can support defensible, high-quality review outcomes.

Best Practices: Technology Assisted Review

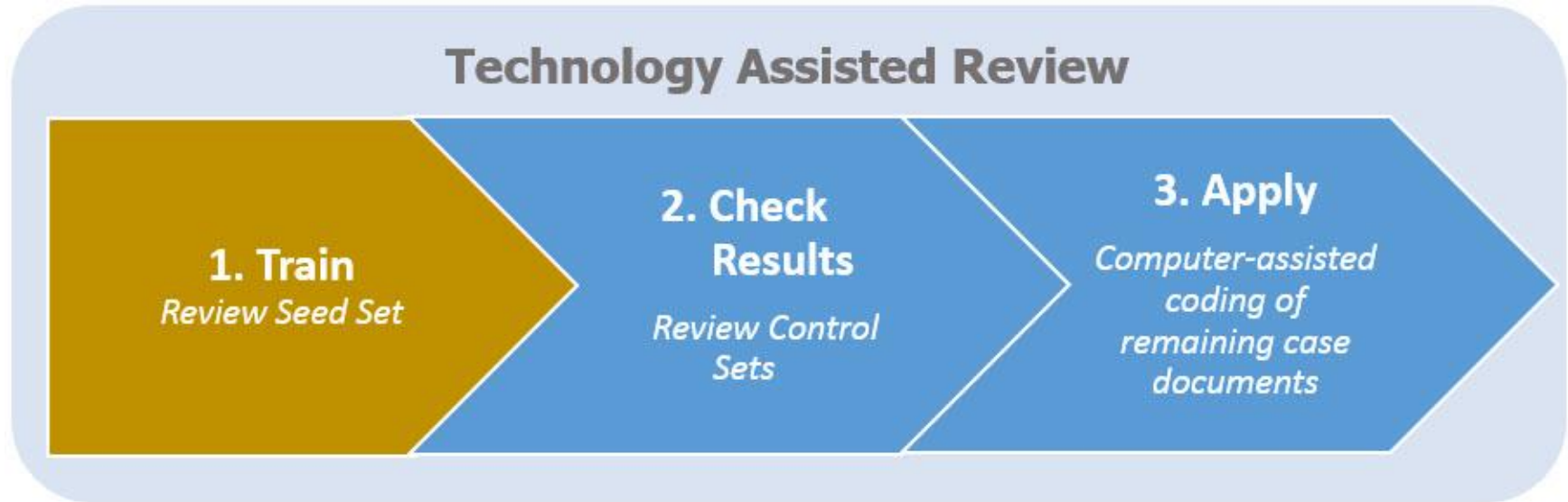
Why Use TAR/Predictive Coding?

CASE STAGE	
Collection	8%
Processing	19%
Review	<u>73%</u>
Total	100%

- Best opportunities for further cost savings will be reducing review costs.
- Technologies and process improvements, like TAR, reduce costs by increasing attorney review efficiencies

Best Practices: Technology Assisted Review

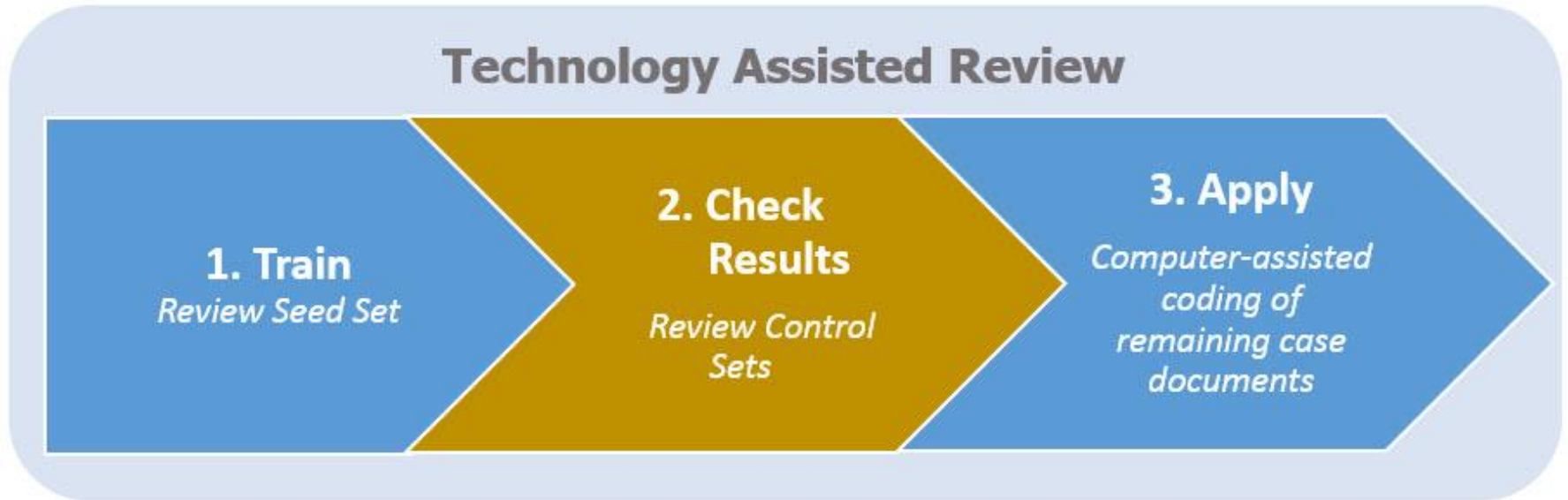
How Does TAR/Predictive Coding Work?



- A randomized sample of $\sim 2,400$ documents, a **seed set**, is selected from the collection.
- A skilled document review professional reviews and codes the **seed set**.
- The coding decisions made in reviewing the **seed set** train the predictive coding algorithm to identify responsive content in the remaining documents.

Best Practices: Technology Assisted Review

How Does TAR/Predictive Coding Work?



- Iterative samples of 25 computer-reviewed documents, **control sets**, are inspected for coding algorithm accuracy.
- The responsiveness designation assigned to the document by the computer is either confirmed or overturned.
- An **F-score** - derived from **precision** and **recall** measures - indicates the stability of the TAR results.

Best Practices: Technology Assisted Review

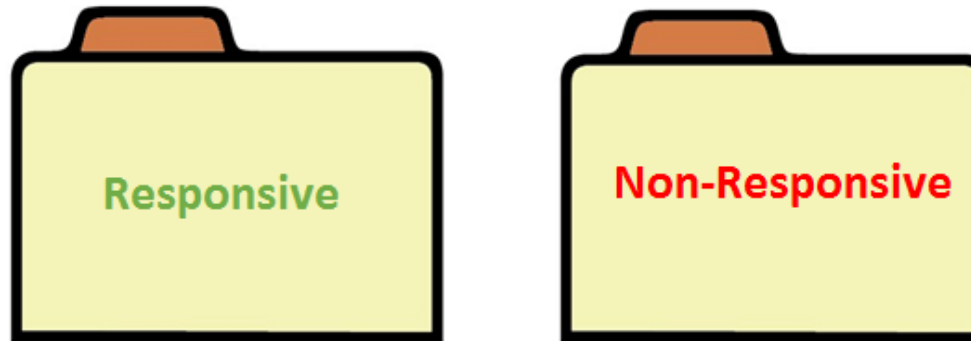
How Does TAR/Predictive Coding Work?



- The TAR algorithm reviews the document collection based on how it was trained during **seed set** coding and **control set** review.
- Remaining Documents are tagged as responsive/non-responsive.
- The speed at which the document collection is reviewed by the TAR algorithm is largely based on the computing resources applied to the task.

Best Practices: Technology Assisted Review

Understanding TAR/Predictive Coding Results



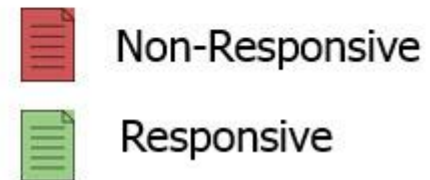
TAR/Predictive Coding results (**F-scores**) indicate:

- What proportion of the responsive documents were found by the algorithm within a particular margin of error (**recall**)
- What percentage of documents marked responsive are *actually* responsive within a particular margin of error (**precision**)

Best Practices: Technology Assisted Review

Understanding Results: Precision & Recall

High Recall, High Precision: All of the responsive documents in the collection were appropriately coded by the algorithm (high recall). All of the documents produced are actually responsive (high precision). Best possible outcome.



Best Practices: Technology Assisted Review

Understanding Results: Precision & Recall

Precision: A measure of how often the algorithm accurately predicts a document to be responsive; the percentage of produced documents that are actually responsive.

Recall: A measure of what percentage of the responsive documents in a data set have been classified correctly by the algorithm.

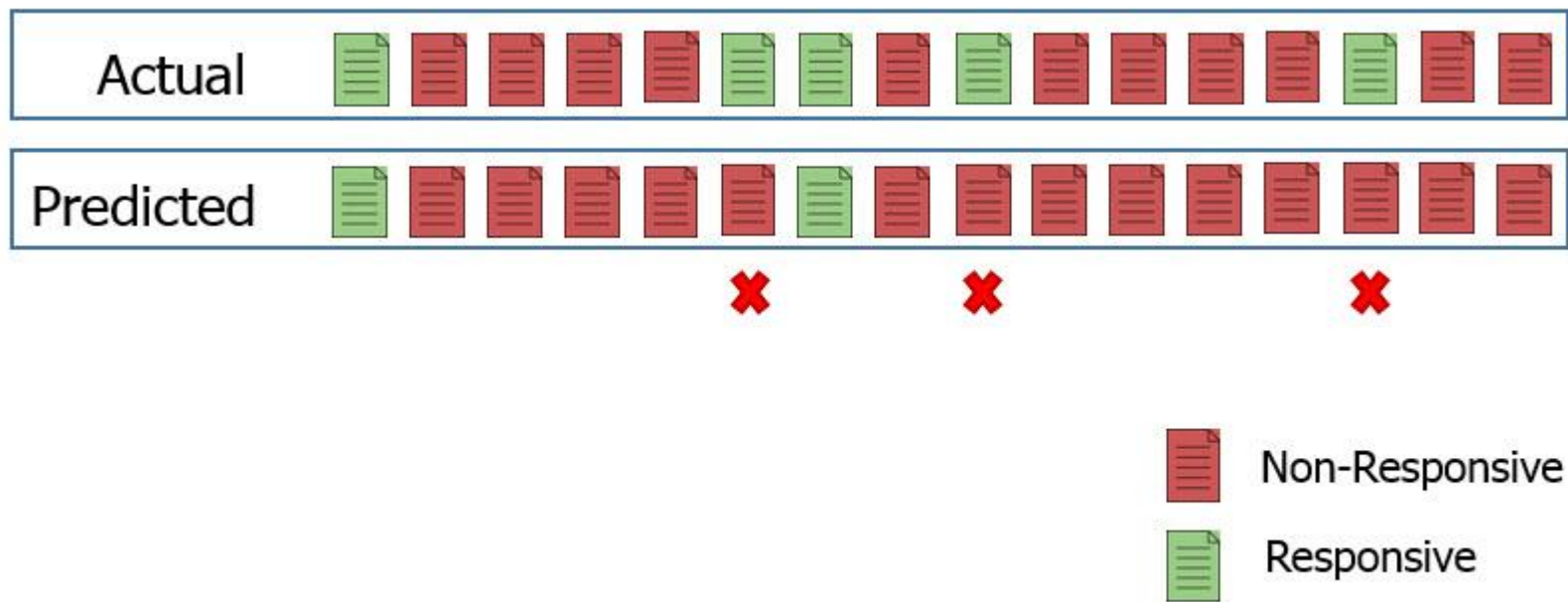
F-Score: Harmonic mean of **precision** and **recall**.

****Note:** F1 scores should not to be interpreted as a measure of review quality but rather as an indication of **1)** how well the case lends itself to TAR and **2)** the quality of the seed set training.

Best Practices: Technology Assisted Review

Understanding Results: Precision & Recall

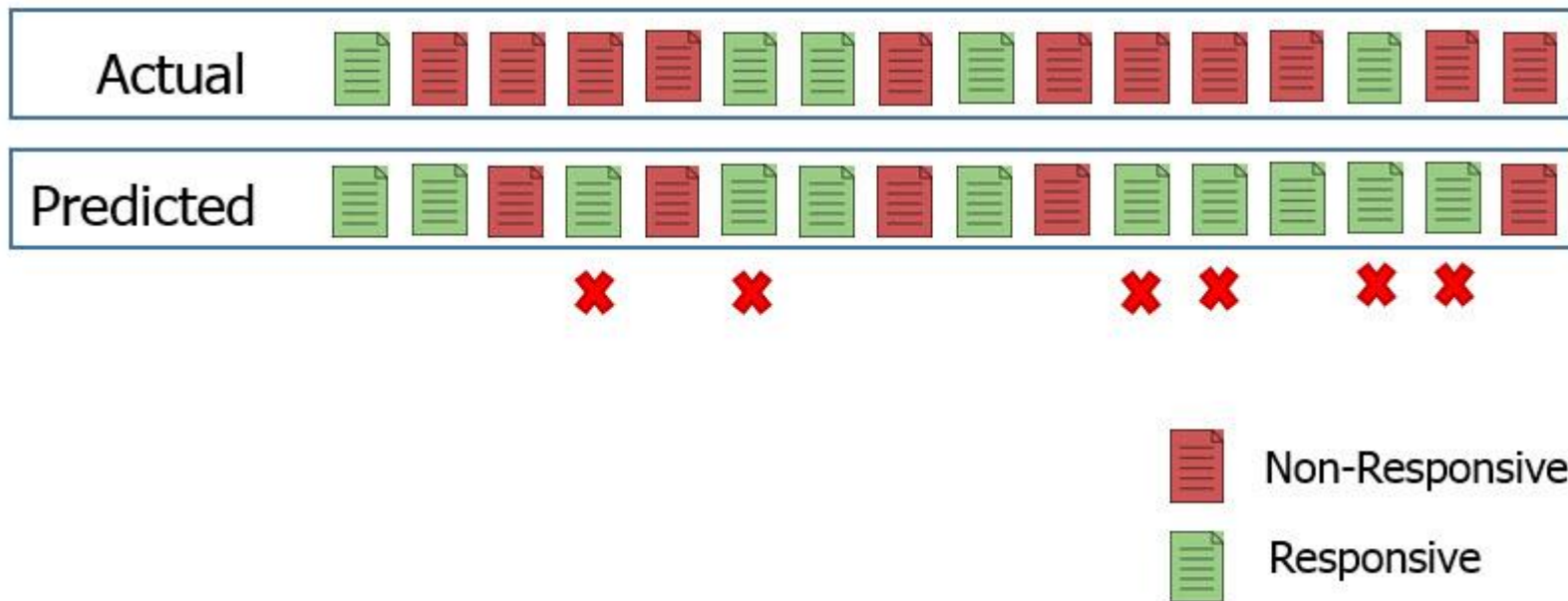
Low Recall, High Precision: Many of the responsive documents in the collection were not appropriately coded by the algorithm (low recall). However, a high percentage of the documents produced are responsive (high precision). Increased risk of under-producing.



Best Practices: Technology Assisted Review

Understanding Results: Precision & Recall

High Recall, Low Precision: All of the responsive documents in the collection have been appropriately tagged by the algorithm (high recall). However, many erroneous documents were incorrectly marked responsive (low precision).



Best Practices: Technology Assisted Review

Comparing Outcomes: TAR v. Manual Review

From the **Sedona Conference** Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery:

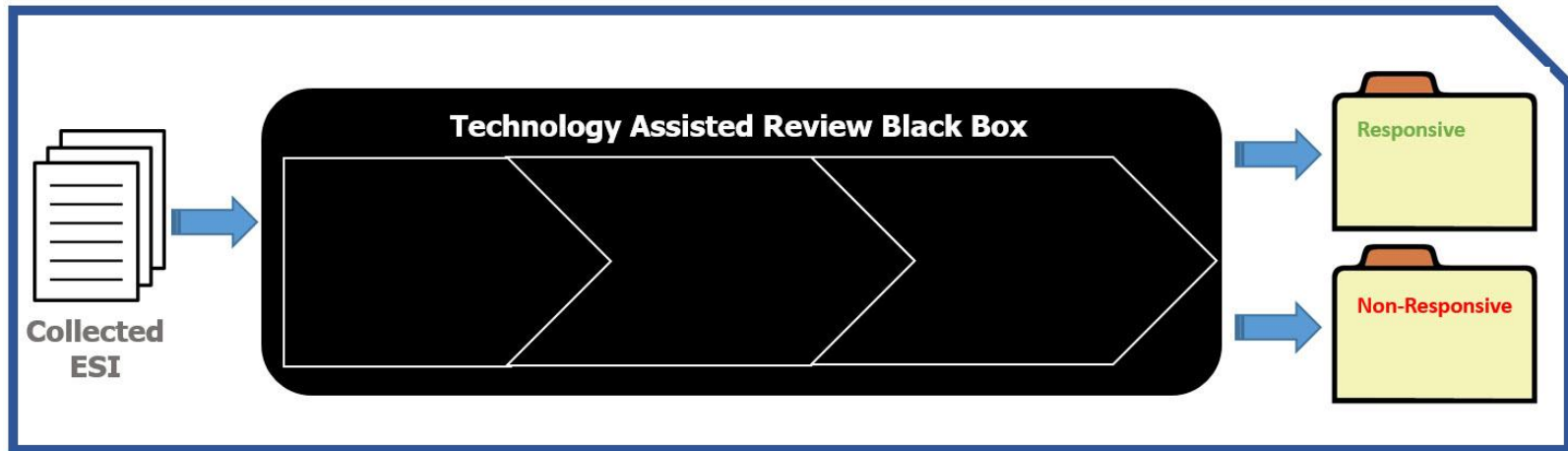
"[T]here appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible ... Even assuming that the profession had the time and resources to continue to conduct manual review of massive sets of electronic data sets (which it does not), the relative efficacy of that approach versus utilizing newly developed automated methods of review remains very much open to debate." (2007)

From the **TREC** (Text Retrieval Conference) Legal Track:

"Overall, the myth that exhaustive manual review is the most effective – and therefore, the most defensible – approach to document review is strongly refuted. Technology-assisted review can (and does) yield more accurate results than exhaustive manual review, with much lower effort...Future work may address which technology-assisted review process(es) will improve most on manual review, not whether technology assisted review can improve on manual review." (2009)

Best Practices: Technology Assisted Review

The Importance of Transparency

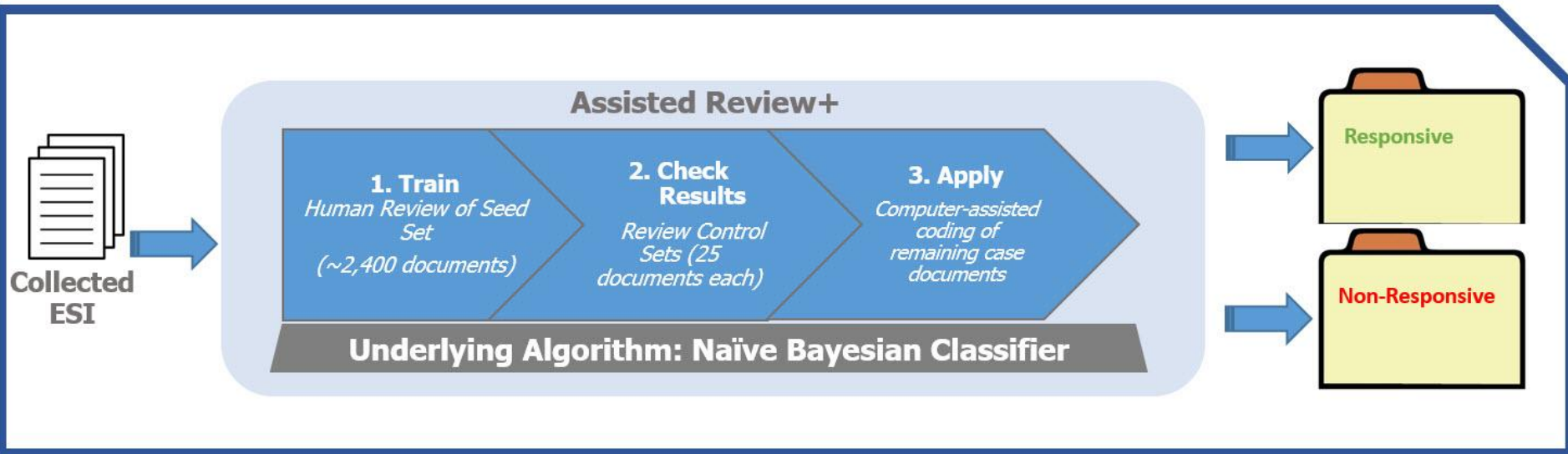


Defensibility: Without understanding how a particular TAR/predictive coding methodology works, it becomes difficult to explain *why* the algorithm made certain coding decisions.

TAR is No Panacea: TAR is not meant to be used in any and all review situations. Without understanding how a particular TAR/predictive coding methodology works, it is impossible to determine if it is appropriate for your case.

Best Practices: Technology Assisted Review

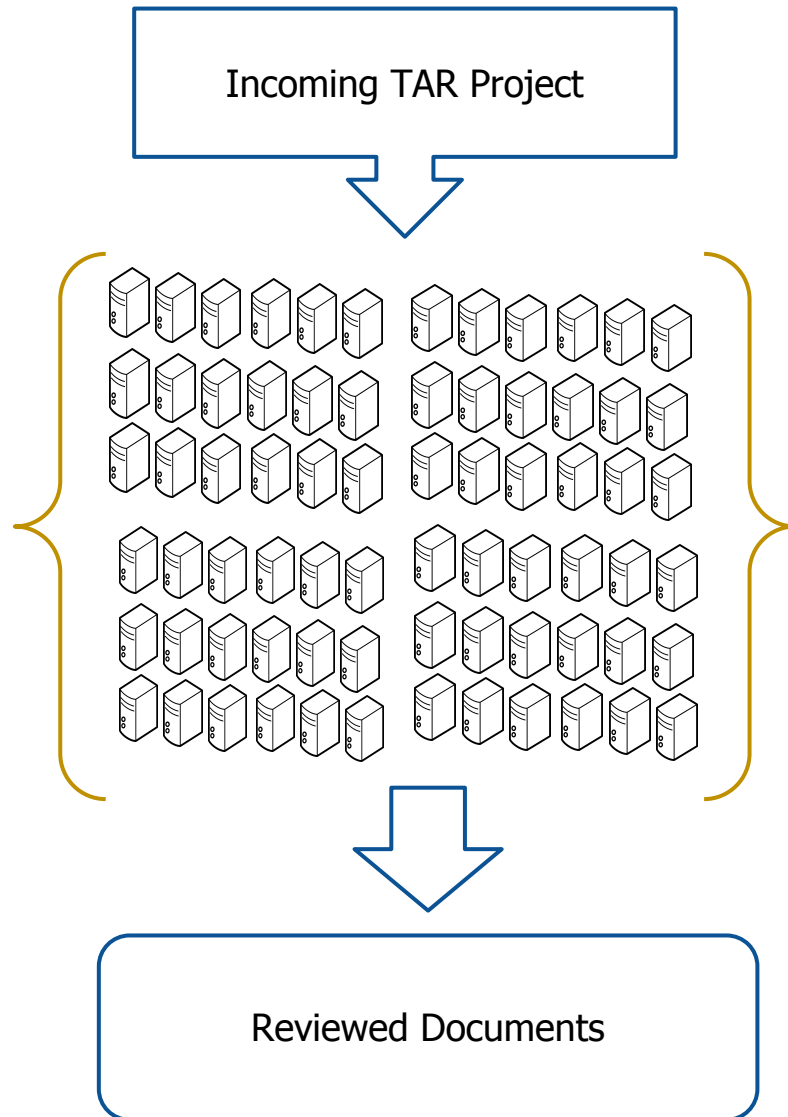
The Importance of Transparency: Assisted Review +



- In TAR, Bayesian Probability models the likelihood of something being true about a document, i.e. responsive, based on the millions of data connections created while training the seed set.
- A Naive Bayesian Classifier, used in Assisted Review+, is a probability model with assumptions that allow for pattern recognition among multiple independent variables.

Best Practices: Technology Assisted Review

The Importance of Scalability



- Applying more server resources to a TAR/predictive coding task will increase throughput.
- TAR offers an exponentially faster workflow compared to manual review. Leveraging scalable architectures maximizes the value of this benefit.

Best Practices: Technology Assisted Review

Summary

- TAR/Predictive Coding allows a skilled reviewer to train a computer algorithm to identify responsive and non-responsive documents .
- You can use TAR/Predictive Coding to increase review speed, decrease review costs, and improve the quality of review results
- TAR works by teaching a seed set, testing the algorithm against control sets, and applying the improved algorithm to the remainder of the collection
- Predictive coding performance results are communicated in the form of precision and recall scores
- It is important to know the underlying logic of the TAR algorithm to interpret, explain, and defend your results.
- Scalable, transparent predictive coding workflows maximize the intended benefits of technology assisted review.

Thank You

Contact Info

Karsten Weber:

karsten@lexbe.com

(512) 686-3382

Stu Van Dusen:

svandusen@lexbe.com

(512) 843-7672

Webinar Questions:

webinars@lexbe.com

www.lexbe.com/assisted-review

**eDiscovery
Webinar Series**

lexbeSM