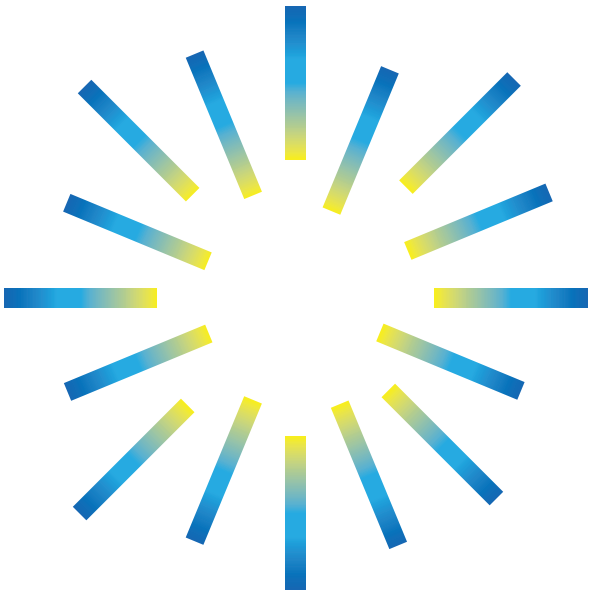

FRONTEO

AI-Enhanced Review with KIBIT

Lilith Bat-Leah
Director, Data Science

Sayaka Nishino
AI Review Manager

Vladislav Lobatchev
Director, Data Science
Engagement Management

Ryo Shiba
Review Manager

Hideki Takeda
Chief Technology Officer

FRONTEO, Inc

FRONTEO USA, Inc.
777 Third Ave, Fl 17 (US Headquarters)
New York, NY 10017
Phone: (866) 803 7668
Email: usinfo@fronteo.com

fronteousa.com

New York | Washington DC | San Francisco |
Los Angeles | Denver | London | Seoul |
Taipei | Tokyo

AI-Enhanced Review with KIBIT

Lilith Bat-Leah
Director, Data Science

Sayaka Nishino
AI Review Manager

Vladislav Lobatchev
Director, Data Science
Engagement Management

Ryo Shiba
Review Manager

Hideki Takeda
Chief Technology Officer

FRONTEO, Inc

Table of Contents

Executive Summary 4

1. Document Review Challenges in the Era of Exponential Data Growth 4
2. Advances in Artificial Intelligence Make Efficient, High-Quality Review Possible 4
   - 2.1 How AI-Enhanced Review Works 4
   - 2.2 Benefits of AI-Enhanced Review: Speed 6
   - 2.3 Benefits of AI-Enhanced Review: Enhanced Quality Control 7
3. Specific Approaches for the PoC 8
   - 3.1 Comparing AI-Enhanced Review to Linear Review 8
4. The Effectiveness of AI-Enhanced Review from PoC Results 10
5. Conclusion 11
6. Acknowledgment 11

Executive Summary

FRONTEO has been using artificial intelligence (AI) for several years, not only within eDiscovery review workflows, but also within organizations — ranging from finance to healthcare to government — looking to cut through voluminous data to find the information required to improve the efficiency of their operations. Seeing an opportunity to do the same for legal document review, FRONTEO inserted KIBIT, its proprietary AI engine, at the heart of its analytics review, to deliver the advantages of AI-Enhanced review to clients. Compared with most other tools on the market, KIBIT requires only a small amount of training data to maximize review efficiency.

To intelligently integrate KIBIT into document review, FRONTEO developed an AI-Enhanced review workflow that maintains review quality while increasing overall review speed. To achieve this, KIBIT compares review target documents against relevant and non-relevant training documents, assigns each document a relevancy score between 0 and 10,000, and then reorganizes and groups the review target documents according to their relevancy score. With KIBIT, FRONTEO creates a defensible AI scoring threshold to substantially reduce the number of documents that require manual review. KIBIT also improves the effectiveness of manual review by prioritizing probable relevant documents by score for batching.

In addition to AI-powered scoring of documents, KIBIT offers a suite of reporting functions and analytics that makes the review process transparent and defensible while enhancing review quality control. The combination of KIBIT with FRONTEO's AI-Enhanced review workflow nearly doubles the review speed of standard linear review and substantially reduces the number of documents that require attorney review.

1. Document Review Challenges in an Era of Exponential Data Growth

Litigators face new challenges as the volume of electronic data continues to grow exponentially. Where it was once common practice for a case team to personally review every document produced in a proceeding, today many litigators must rely on the combination of technology and a staff of attorney reviewers to prepare their clients' document productions. This reality is accompanied by myriad risks and inefficiencies. Voluminous data repositories are rarely whittled down appropriately by search parameters alone. Additionally, human reviewers sometimes misunderstand instructions or lapse into carelessness. These complications force case teams to rely on highly variable results to slowly surface the most important evidence in the case. Attorneys are forced to forego quality, cost-savings, or speed as they wrestle with the discovery challenges posed by big data. Fortunately, the phenomenon causing these issues is also driving solutions.

Best practice solutions include both technological and human strategies. On the technological side, machine learning is deployed for AI classification of relevant and non-relevant documents. Behavioral science expertise builds on that to provide methodical, efficient workflow analysis and training. Taken together, FRONTEO has seen measurable improvement in both speed and accuracy over traditional linear review.

Jack Fornaciari, a BakerHostetler partner with a thriving antitrust practice, is often confronted with tight discovery deadlines requiring thousands of hours of document review. His clients' data is often nuanced and sensitive, so review teams must consider each document carefully. To demonstrate that advances in technology can successfully reduce his clients' litigation spend, FRONTEO provided an AI-Enhanced review Proof of Concept (PoC) to Mr. Fornaciari.

2. Advances in Artificial Intelligence Make Efficient, High-Quality Review Possible

2.1 How AI-Enhanced Review Works

FRONTEO's KIBIT empowers attorneys to harness the power of AI throughout the review process for improved transparency, defensibility, and efficiency. While the term "AI" may

evoke images of a post-apocalyptic future dictated by sentient machines, countless business leaders and scientists have published work emphasizing the role of AI in enhancing the product of human intelligence rather than replacing it. As with any technology, AI is merely a tool that allows us to accomplish more with less.

AI-Enhanced review is the next step in the Technology Assisted Review (TAR) arena, which is already widely accepted in eDiscovery. Where many review platforms offer a TAR feature, FRONTEO extends the utility of their offerings to managed review. FRONTEO leverages KIBIT, its AI engine, to make every stage of the review process more efficient and effective. To prove the accuracy of KIBIT, FRONTEO produced simulated results using actual documents and review designations from one of Mr. Fornaciari's prior cases. FRONTEO then staffed and conducted a full review using the documents from the case and the workflow outlined below to demonstrate how KIBIT increases both human review speeds and the efficiency of quality control throughout the duration of a live document review project.

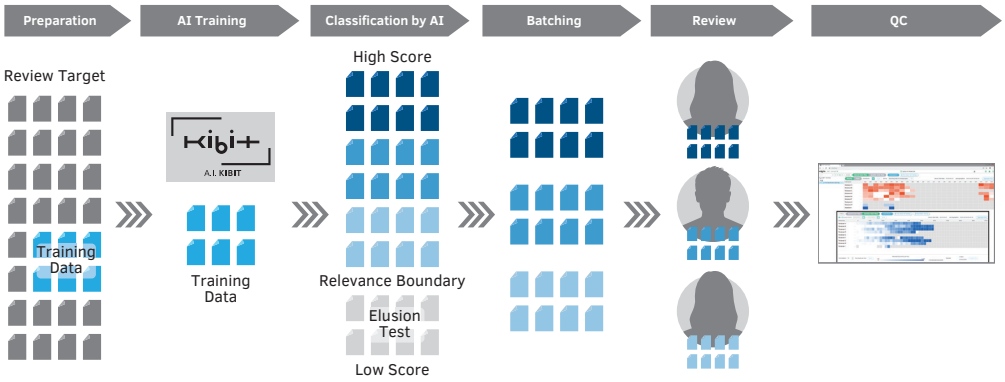

Figure 1: AI-Enhanced Review Process

Figure 1 illustrates the AI-Enhanced review workflow used in the PoC. First, the target review population was established. For any given matter, the entire document corpus can be designated for review, or the review target can be narrowed. Targeting is frequently achieved through a keyword search or metadata parameters. In some cases, unsupervised machine learning features like concept analysis or clustering are used to narrow the review target. The target review population used in the PoC came from approximately 40 custodians and was culled down to nearly 11,550 documents using search term, date, file type, and text size filters. FRONTEO's AI engine, KIBIT, was then trained using a sample of the review target. A subset of 1,356 documents was reviewed and coded to comprise the KIBIT training data set.

FRONTEO's proprietary AI engine, KIBIT is trained on a sample, and each document is assigned a score according to relevance. This relevancy scoring informs review workflow throughout the project by prioritizing documents for attorney review, eliminating non-relevant documents from human review, and enhancing QC processes. KIBIT's resulting predictive model is evaluated using typical statistical indicators such as recall, precision, and elusion rates. Once the client/law firm is confident in KIBIT's classification of documents, KIBIT scoring is applied to the entire review target.

Documents below a scoring threshold are sorted into a non-relevant discard set and excluded from human review. Likely relevant documents are batched by KIBIT so that documents most likely to be relevant are prioritized for review. The scoring threshold for the discard set is determined on a case-by-case basis, depending on the case team's review goals and the statistical model associated with that case. Estimated recall and precision metrics given by the scoring threshold guide attorneys through proportionality and defensibility considerations. To confirm that the scoring threshold is appropriate, an elusion sample is reviewed to verify that the AI-reviewed documents are non-relevant. Scoring and relevancy distributions are also used to enhance QC workflows.

FRONTEO measured true recall, precision, and elusion rates in the simulation phase of the PoC to demonstrate KIBIT's accuracy. Coding designations from the real case showed that KIBIT would have reduced the population for human review to 36% of the original data volume while achieving a 75% true rate of recall.

Item		Numbers
Training data	Statistical random sampling • Confidence level 95% • Margin of Error +/- 2.5%	1,356
	Relevant in training data	234
	Richness	17.30%
Recall Rate (True value)		75.00%
Precision Rate (True value)		35.20%
Elusion Rate (True value)		6.60%
Total document volume (Approximate)		11,550
AI-Only Review (=Non-responsive docs)		64.0%

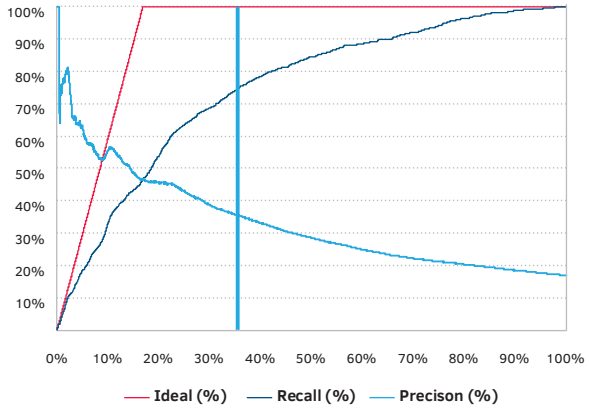

Figure 2: Quality Metrics

Like all AI-powered solutions, KIBIT is designed to augment human effort. This is accomplished not only by propagating attorneys' decisions to similar documents, but also by facilitating an optimized workflow that accelerates review speeds and flags potential mistakes. To demonstrate enhancements, FRONTEO fully replicated the human review component of Mr. Fornaciari's matter. As on a real matter, statistical estimates of recall, precision and elusion rates were used to inform the workflow in this phase of the PoC.

2.2 Benefits of AI-Enhanced Review: Speed

The pace of document review tends to increase when similar documents are reviewed together. Attorneys can breeze through batches when coding designations are the same for dozens of documents in a row. With AI-Enhanced review, we use KIBIT's relevancy score to create high- and low-scoring batches.

Average review speeds tended to be slower, around 60 documents per hour, for high-scoring batches that contained mostly relevant documents. In contrast, low-scoring batches contained a greater proportion of non-relevant documents and were therefore reviewed at a faster pace, around 90 document per hour. *Figure 3* illustrates the change in review speed with scored batching using KIBIT, compared to the linear review speed of the same documents.

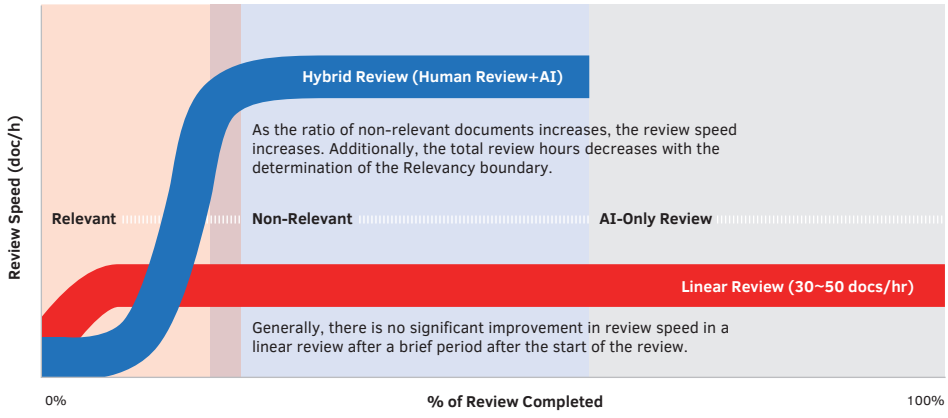

Figure 3: Review Speed. Hybrid Review vs. Linear Review

2.3 Benefits of AI-Enhanced Review: Enhanced Quality Control

Beyond the significant first-pass review efficiency gains afforded by KIBIT's relevancy scoring, FRONTEO also streamlines second-pass review with targeted quality control (QC). A "QC Heat Map" gives review managers a visual representation of anomalies in reviewer coding. In this AI-driven feature, the results of human reviewers are measured against the predicted classification of the KIBIT score. The results of the QC Heat Map can be used to evaluate the efficacy and quality of the human reviewer. The QC Heat Map and the use of its feedback improves reviewer training and extends the usefulness of AI into the review management process.

The QC Heat Map compares the AI-modeled prediction assigned to a document with a human reviewer. The results are compared using visual color-coding. The QC Heat Map's color diagram empowers the review manager to identify where human reviewers and artificial intelligence disagree about the classification or coding of a document. Quality control can be efficiently and effectively managed with the assistance of this technology.

In the *QC Heat Map* diagram, each cell represents a set of documents within a range of KIBIT relevancy scores. The cell color is determined by the percentage of documents a reviewer has coded as relevant. If most or all of the documents are coded non-relevant, then the cell is light blue or white. If most or all of the documents are coded relevant, then the cell is deep blue. If the reviewer has coded the documents correctly, the Heat Map will maintain a clean gradient from white to blue.

If the reviewer's coding does not create a clean gradient, and the cell colors appear disrupted, it is a visual signal to check on those documents to see if there are any coding mistakes. See the below Heat Map as an example.

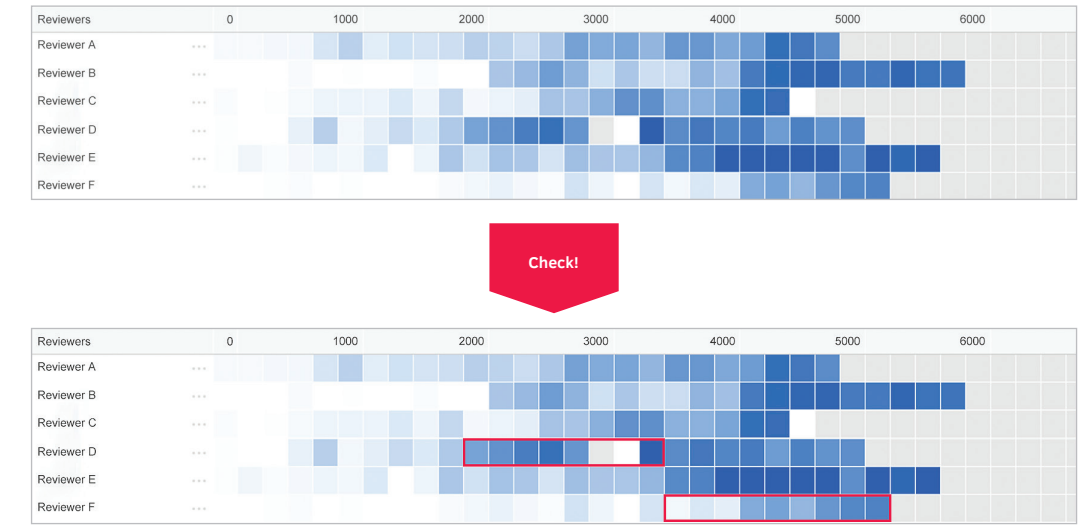

Figure 4: QC Heat Map

In *Figure 4*, Reviewer F has coded several documents with a high-relevancy score as non-relevant. In contrast, Reviewer D has coded several documents with a low-relevancy score as relevant. The QC Heat Map instantly reveals that their results are anomalous and differ from other reviewers. The discrepancy indicates the likelihood of a coding error, attributable to a misinterpretation of documents or a misunderstanding of review instructions.

As part of this PoC. we compared a QC by Heat Map to a 10% random QC and found that when the review manager used the KIBIT Automator Heat Map, overturns were found much more efficiently, resulting in much less time spent on QC.

Item		Value
Heat Map	QC Document volume	185
	Overall Overturn rate (Docs)	58.9% (114/185)
10% Relevant QC 5% Non-Relevant QC	QC Document volume	1068
	Overall Overturn rate (Docs)	18.5% (197/1068)

Figure 5: Comparing the Efficiency Between Heat Map and 10% QC

The QC Heat Map provides an overview of the quality of reviewer coding. Potential issues are flagged for review managers, giving them insight that could not be replicated through traditional keyword or tag searches used in most review quality control checks.

Another review management function in KIBIT Automator is the Speed Heat Map, which provides a visualization of each reviewer's pace. Reviewers moving too slowly can be a problem on an urgent review. On the other hand, an abnormally fast pace may be indicative of careless review and classification. By checking the Speed Heat Map hourly or daily, the review manager can check review speeds at a glance and determine who is going too fast or too slow to better manage the review.

For example, looking at the *Speed Heat Map* diagram below, you can see that Reviewer D was working at an unusually slow rate during 04h. By referencing the Speed Heat Map, the review manager was able to immediately ask the reviewer if they encountered a problem document during that hour.

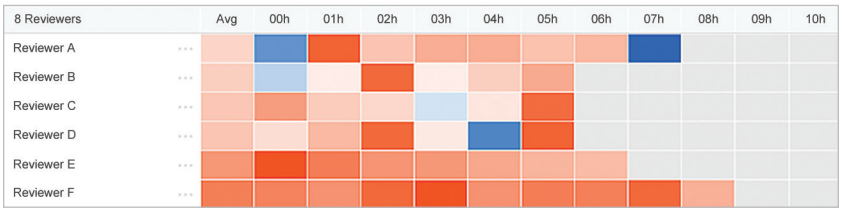

Figure 6: Speed Heat Map

3. Specific Approaches for the PoC

3.1 Comparing AI Review to Linear Review

FRONTEO controlled for several variables to prove the efficiency of AI-Enhanced review as compared with linear review. Variability in the data was controlled by using the exact same documents that were reviewed originally so that outcomes could not be attributed to differences in the length or nature of the documents. To control the workflow, FRONTEO grouped document "families" together, a common practice in linear review. Additionally, reviewers were required to code the same fields as in the original review to control for the complexity of the review.

FRONTEO engaged the review manager who managed the original linear review to lead a team of six reviewers for the PoC. Because a familiarity with the documents would give the reviewers an advantage not attributable to KIBIT Automator, none of the six reviewers were among those who had been assigned to the original case. All six possessed comparable review skills in order to obtain unprejudiced review speed results. AI metrics were applied for relevance only, but the reviewers selected designations for the same number of fields (including issues, confidentiality, and privilege) using a coding panel identical to the one used in the original review.

Training data was also re-reviewed by the review manager who followed the same protocol as in the original case. Notably however, in the original review, the protocol required many updates while the matter was ongoing. The result was a final coding protocol inconsistent with the one used at the outset of the project. While a sophisticated workflow would allow FRONTEO to address this in a live matter, the review manager re-reviewed the 1,356 documents comprising the training data according to the guidelines specified in the original protocol for the sake of clarity and simplicity. This rendered the manual review and the AI decision consistent. For the elusion test and cut-off, the review manager re-reviewed 380 documents. An elusion rate of 4.47% validated classification of non-relevant documents with no human review. We call this “AI-Only Review”.

AI-Enhanced review gives reviewers the advantage of AI highlighted insights on their review screen. When a document’s score is higher than a certain threshold, KIBIT Automator will make a tag recommendation to show reviewers documents with highly probable characteristics of relevance. The other AI function highlights sentences most likely to contain relevant information. KIBIT identifies sentences in highly probable relevant documents from training data, and highlights them automatically to help reviewers quickly determine document relevance. Using these functions, and with batches prioritized from high score to low, we saw a dramatic increase in overall review speed, which we call “Hybrid Review”, a combination of human and AI review.

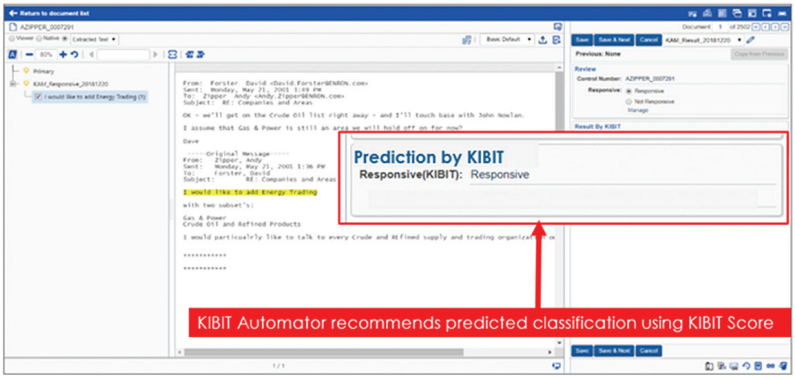

Figure 7: Tag Recommendations

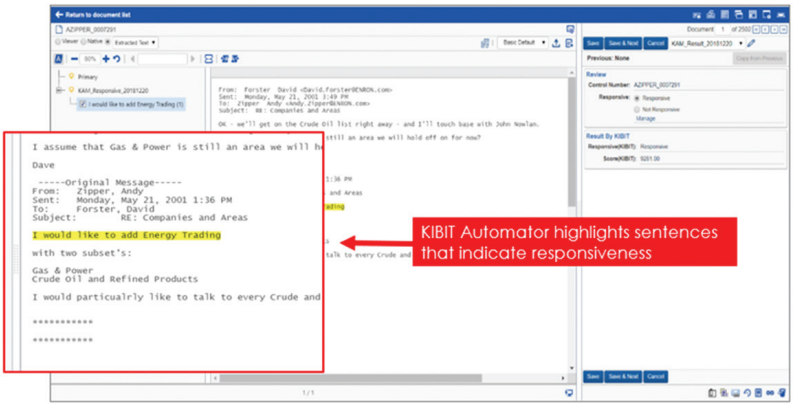

Figure 8: Sentence Highlighting

FRONTEO evaluated the PoC results via a number of relevant metrics and found that increasing the speed of review significantly reduced the overall cost of the project. Equally important, AI-Enhanced review allowed for more accurate review of large data volumes within delineated timeframes. Both results lead to tangible, consequent benefits for the client.

4. The Effectiveness of AI-Enhanced Review from PoC Results

Comparing the results, hybrid review yielded a pace of 73.7 documents per hour as opposed to 40 documents per hour with linear review. In combination with AI-Only review, total review time was cut by 48%, meaning AI-Enhanced review was nearly 2 times faster than linear review.

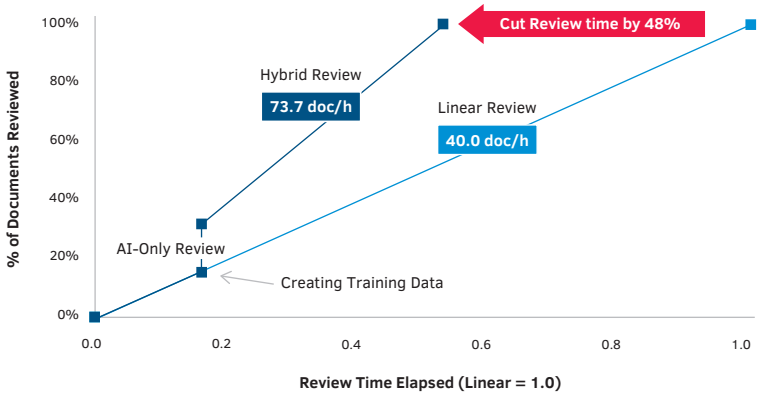

Figure 9: Comparing Review Speed of Hybrid and Linear Review

Notably, as seen in *Figure 10*, prioritized batching allowed reviewers to identify more than 80% of the relevant documents in the first 3 days of the review.

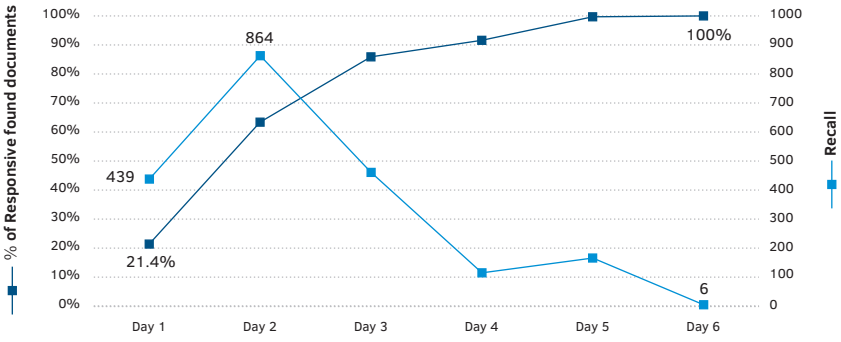

Figure 10: Efficiency of Finding Relevant Documents

As hypothesized, review speed increased as batch richness fell. Reviewers were able to fly through the final batches due to the high percentage of non-relevant documents they contained. *Figure 11* demonstrates these results.

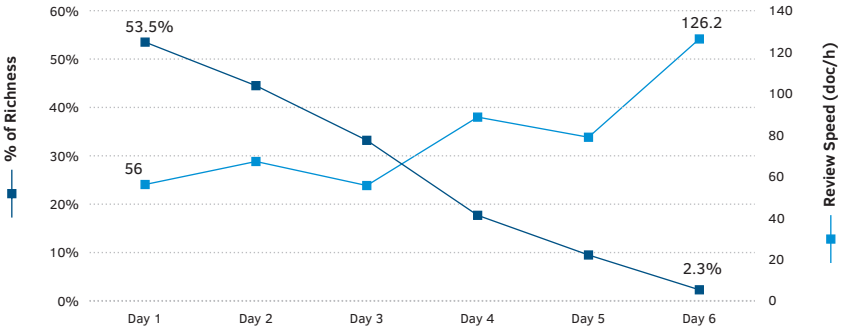

Figure 11: Transition on Richness and Review Speed

Finally, by using FRONTEO's KIBIT, FRONTEO reduced the number of documents for review by 64% and achieved an average review speed of 91.6 documents per hour when accounting for AI-Only review, cutting review time by 48% compared to the original linear review. The PoC confirmed that strategic deployment of artificial intelligence will result in ultra-efficient reviews without sacrificing defensibility. "We substantially shortened the review timeline by integrating FRONTEO's KIBIT without sacrificing quality," said Jack Fornaciari at BakerHostetler. "The results confirm the efficiency of using enhanced AI in document review and we look forward to utilizing it in future cases."

5. Conclusion

These results are very promising. KIBIT aids human effort in a quantifiable way, resulting in significant benefits to FRONTEO's clients. In particular, KIBIT Automator is able to analyze much larger volumes of data with the same small amount of training data as in the PoC exercise. Human error is greatly reduced, while speed and accuracy are amplified by KIBIT's AI-Enhanced review. The speedy, accurate analysis leads to greater efficiency for all stake-holders in a legal action, from defense teams to opposing counsel and judiciary. We all have a stake in increasing the efficiency while lowering the cost of litigation, and that promise is being made real with AI-Enhanced document review.

Moreover, FRONTEO is delivering even better, more rewarding results in live cases where project managers and data scientists are able to optimize database configuration and workflow around the specific details of the project and review. FRONTEO's approach combines human skills and experience with cutting-edge technology to meet client needs more efficiently and effectively. FRONTEO's data scientists are able to tailor implementation and workflow to the needs and specifics of each client's matter, no matter the scale. In fact, KIBIT's abilities are best tested when it is applied to analyzing and reviewing dauntingly large data sets. FRONTEO leverages both human and artificial intelligence to deliver solutions specifically suited to the needs and goals of our clients.

6. Acknowledgment

We would like to thank Jack Fornaciari at BakerHostetler for his feedback and his collaboration on this PoC. We would also like to thank the team of experts at FRONTEO for their cooperation on carrying this PoC, and their help on editing this paper.

FRONTEO USA, Inc.
777 Third Ave, Fl 17 (US Headquarters)
New York, NY 10017
Phone: (866) 803 7668
Email: usinfo@fronteo.com

fronteousa.com

New York | Washington DC | San Francisco | Los Angeles | Denver | London | Seoul | Taipei | Tokyo

FRONTEO

About FRONTEO.

Everything we do comes from our service-oriented culture that puts the client at its center. We develop and constantly improve leading-edge technology. And we market custom, seamless services that create value for our clients, employees, consumers and shareholders. Our name is no coincidence. “FRONTEO” helps us focus while looking forward, empowering us to apply our AI technology to the mission-critical legal market. We combine our machine learning with unparalleled attention to our clients. At FRONTEO, *you* are the center of our universe. You define your needs, and we are your helpmates and facilitators.

Together, now and in the future, we succeed.