

# Lighthouse Review Achieves Unprecedented Accuracy and Efficiency in Federal Criminal Investigation

---

CASE STUDY



# Lighthouse responsive review paired with proprietary linguistic modeling and Key Document Identification (KDI) services drastically reduced the need for manual review in an extremely complicated dataset.

## What they needed

A global transportation company was under investigation for possible infractions of the Foreign Corrupt Practices Act (FCPA) in India. The company's legal counsel needed to quickly produce responsive documents and find key documents to prepare their defense.

However, due to the complicated nature of the dataset, they faced significant hurdles in producing the documents and preparing their defense. First, because of the sensitive nature of the documents collected from executives, it was critical to ensure productions were limited to only documents that related to the India market – something that would be impossible for most technology assisted review (TAR) tools to achieve. Second, counsel was finding it challenging to identify key documents, as bad actors tend to hide incriminating communications.

The company's outside counsel came to Lighthouse because they knew our innovative TAR process, linguistic expertise, and proprietary Key Document Identification (KDI) could overcome these challenges with time- and cost-efficiency.

## How we did it

Lighthouse began work on an initial population of 2M documents, collected from senior executives located in both the US and India. To ensure that the final production set was limited to documents related to the India market, Lighthouse used a hybrid approach, layering machine learning (ML) TAR workflows with linguistic modeling.

First, knowing that machine learning would struggle to distinguish between discussions that were based in the US versus India, particularly when those discussions co-occurred frequently within the same documents, H5 created a linguistic model and adjusted the machine learning population to exclude documents that contained common domestic references but lacked common India references. By doing this, we could ensure that our machine learning algorithms would not be hobbled by "mixed signals" regarding responsiveness, and could instead effectively detect and generalize patterns of relevant language over the substantial portion of the population that was free of domestic discussions.

The Lighthouse responsive review team developed geographic filters based on documents that were confirmed as India-specific. The team then trained ML models with those filters and applied more precise linguistics models to the full

# \$1.7M

TOTAL SAVINGS  
WITH LIGHTHOUSE

2M document population. Because Lighthouse TAR tools analyze both the text and metadata of documents, they are more accurate in identifying responsive context than legacy TAR tools.

The team also developed linguistic models to supplement the ML model and weed out any documents that the TAR models may have over-captured. In this process, trained linguists analyzed the nuances of how the English language was used in communications across different regions of India, further improving the responsive capture, while eliminating documents that pertained to nonresponsive content.

Once outside counsel saw the efficacy of the workflow, they asked Lighthouse to work on a second set of documents, which contained an additional 2M documents collected from custodians located in India.

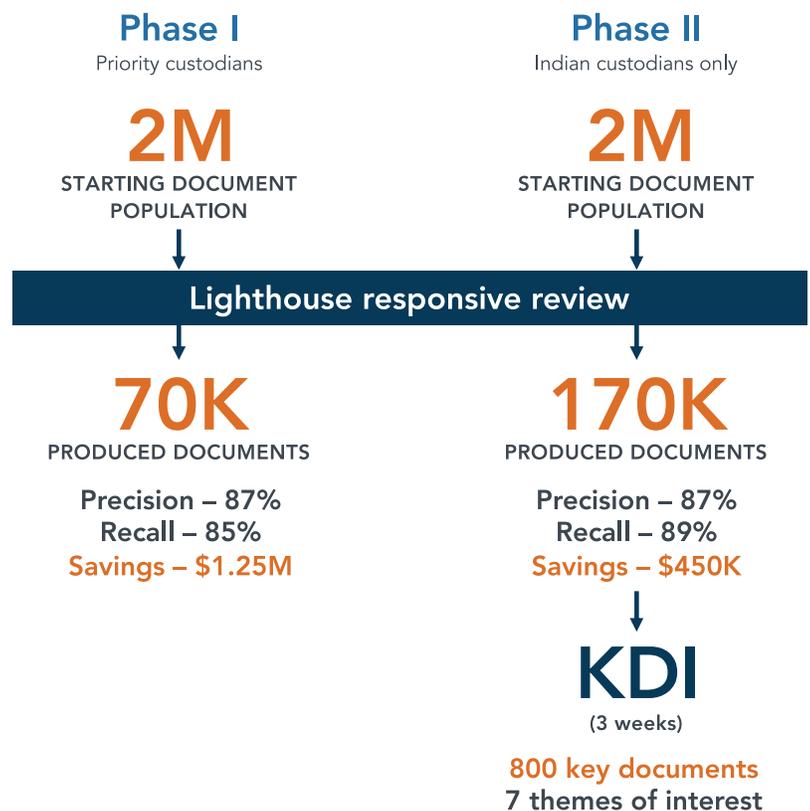
For this phase of the review, Lighthouse leveraged our prior TAR models with updates to account for refined responsiveness criteria and custodial data from different departments and regions within India, to further reduce the responsive set. Then, our KDI team identified evidence critical to the company's defense by focusing in on the language that potential bad actors (vendors, consultants, and other third parties the company engaged with) might use to obfuscate bribery. To do so, the team used the keyword in context (KWIC) feature to target verbs of exchange. KWIC pulled those verbs into a table view so the KDI team could examine what nouns were commonly used on either of the exchange verbs (for example, "give cash" or "provide data") and zero in on the outliers that indicated possible bribery acts performed by outside vendors.

## The results

In the first phase of the review, our proprietary responsive review achieved a precision rate of 87% and a recall rate of 85%. This incredible accuracy successfully narrowed down the production set to the 67K documents pertaining to the India market, out of the initial 2M document population – saving the underlying company \$1.25M.

In the second phase of the review, our TAR workflow achieved a precision rate of 87% and a recall rate of 89%, successfully narrowing down the production set to 170K documents from the initial 2M documents. This workflow saved the client an additional \$450K.

For even more value-add, in just three weeks, Lighthouse KDI uncovered the 810 most important documents out of 250K documents – spanning across seven themes of interest. Leveraging KWIC and linguistic analysis, the KDI team provided valuable investigatory insights that would have otherwise remained purposefully hidden within the data, enabling outside counsel to provide the best defense to the underlying company.





## About Lighthouse

For 25 years, Lighthouse has provided innovative software and services to manage the increasingly complex landscape of enterprise data for compliance and legal teams. Lighthouse leads by developing proprietary technology that integrates with industry-leading third-party software, automating workflows, and creating an easy-to-use, end-to-end platform. Lighthouse also delivers unique proprietary applications and advisory services that are highly valuable for large, complex matters, and a new SaaS platform designed for in-house teams. Whether reacting to incidents like litigation or governmental investigations, or designing programs to proactively minimize the potential for future incidents, Lighthouse partners with multinational industry leaders, top global law firms, and the world's leading software provider as a channel partner.

**Contact us to find out what Lighthouse can do for your business.**

(206) 223-9690 | [lighthouseglobal.com](http://lighthouseglobal.com) | [info@lighthouseglobal.com](mailto:info@lighthouseglobal.com)