DLA PIPER

AN AI CASE STUDY:

Proactive compliance as a service

DLA Piper's clients – and companies broadly – face a common problem across the globe. Every day, vast amounts of unstructured data contain potential hints of future legal problems. The roots of multimillion- or billion-dollar fines, lawsuits, or sanctions live in communications, transcripts, and documents and are often too rare or subtle to assess. Many solutions exist for monitoring structured data, such as receipts and payments, but they often capture problems reactively.

In assessing these challenges, DLA Piper hypothesized that unstructured communications data, documents, transcripts, and other sources could find embers of legal risk – or "needles in a haystack" – before they turned into violations.

But the problem is one of scale. There is often simply too much data for companies to sift through. Human review, while ideal, may be impractical and cost prohibitive. Further, classical statistical sampling approaches are often ill-suited to detect rare events distributed sparsely across millions of documents. The prevalence of problem documents is usually low – often below 0.1 percent of the universe of data. An even greater challenge when monitoring this data is that it is prospective rather than responsive to whistleblowers or known events.

Raising the stakes, regulators increasingly expect organizations to find these needles in a haystack proactively. As John Carlin, Former Acting Deputy Attorney General of the Department of Justice, put it, "[i]t's going to be the expectation here when evaluating compliance programs that corporations are using the same type of analytics to look for and predict misconduct."[1]

Proactive compliance as a service

While DLA Piper believed generative AI was the solution, we also suspected that off-the-shelf generative models would lack the accuracy, nuance, and consistency to work at scale.

After testing available commercial solutions for a client, DLA Piper decided to build and test a tailored approach. We ultimately built a generative AI system for *proactive compliance as a service* (PCaaS) that outperformed our own traditional machine learning (ML) models, as well as a third-party comparator's basic and finetuned generative models and human reviewers using traditional commercial technology-assisted review (TAR).[2] *See* Figure 1.

DLA Piper achieved these results by (i) combining traditional ML and generative AI to leverage their respective strengths, (ii) building small language models (SLMs) trained by the firm's domain-specific lawyers rather than building or relying on large language models (LLMs), and (iii) creating on-demand client-specific models trained on the client's own organizational information, policies, and data in tandem with the firm's legal finetuning. Notably, this process prevents the intermixing of data and models across clients.

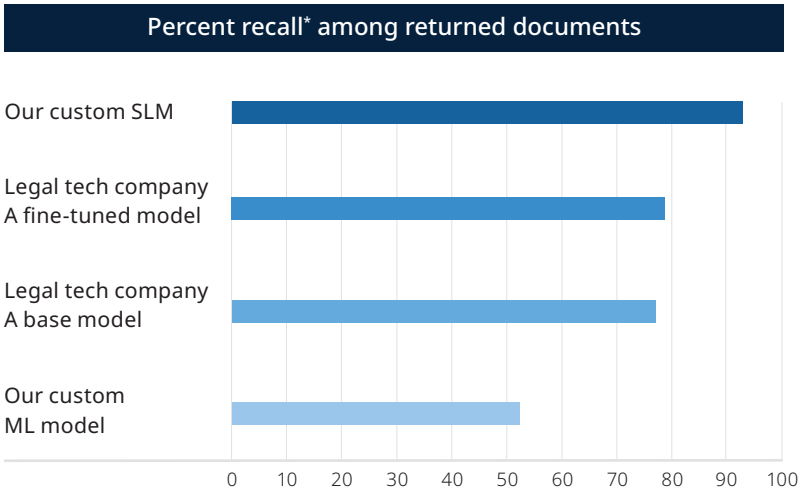


Figure 1: According to our findings, our custom SLM outperformed even other finetuned legal language models offered by other vendors. We expect our SLM recall to increase when our custom ML model is used first to remove irrelevant documents and as we continue to perform ongoing proactive compliance monitoring for clients. *The recall of true positives was validated by human reviewers when looking at the top 8 percent of the same document corpus.

As part of our PCaaS approach, each model is built for a particular client and risk domain (*e.g.,* a given statute, regulation, or issue of concern), to be deployed on an ongoing basis for that purpose. Multiple models can run in tandem to cover different focus areas. Importantly, the AI is hosted by DLA Piper, its performance is managed by DLA Piper, and its results are analyzed by the firm's investigations lawyers so that findings are placed in the proper legal context prior to being communicated to the client. As such, these proactive analyses retain the same level of privilege afforded to after-the-fact reactive investigations.

DLA Piper measured success as a dual issue of scale and proactivity: Can companies cost-effectively use PCaaS to surface (i) needles in a haystack in vast unstructured datasets (scale) and (ii) signals earlier in the chain of events proactively, before embers become fires? Human lawyers established the ground truth by identifying key signals in the data and evaluating the model's performance using precision and recall metrics. We found that our models could both surface early warnings of potential future violations and apply to new regions and time periods.

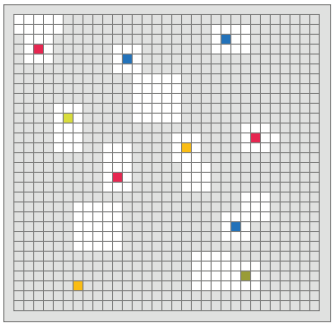
The approach did more than address the scale problem. Our small language model returned two-thirds of critical documents within a return set of only 1 percent of the total document universe. The top 10 percent of documents ranked by DLA Piper's model yielded a recall rate exceeding 80 percent. Not only did these results identify problematic elements, making proactive compliance at scale possible, but they suggested improved quality. *See, e.g.,* illustrative data in Figure 2.

Sample Tiered Model Results

A case study on how training tailored models can imporove cost-effectiveness and outcomes

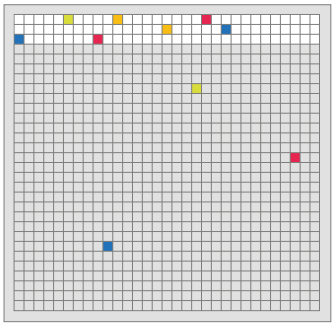
■ Docs removed by ML model ☐ Docs reviewed by attorneys ■ Docs unreviewed ■■■■ Hot docs: each color is a specific topic

Human review $$$
No custom AI models



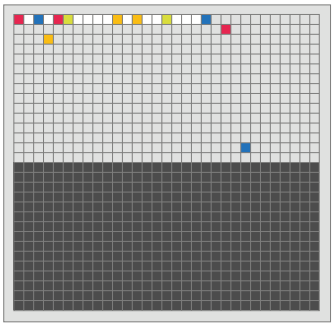
Culled doc volume: 0%
Docs reviewed: 20%
Hot docs identified: 90%

GenAI SLM $$
Custom SLM
(Trained on laws, context, client policies, and lawyer feedback)



Culled doc volume: 90%
Docs reviewed: 10%
Hot docs identified: 70%

ML + GenAI SLM $
Custom SLM + custom ML model
(ML trained on prior data uncovered during legal review)



Culled doc volume: 98.6%
Docs reviewed: 1.4%
Hot docs identified: 70%

Figure 2: These charts show the evolution of the described experiment comparing straight human review to an AI-augmented proactive compliance approach. Our SLM reduced manual review and attorney time significantly but still incurred high inference costs on large volumes of data. The ML pre-filter removed junk and cut inferencing cost and review volume. Combining ML models and SLMs returned approximately 1 percent of the document corpus for human review and still captured key issues. Importantly, our goal is not to locate every single related document, but rather to uncover every substantive issue or problematic circumstance embedded in the data.

Studies of human document review suggest that manual reviewers, even under optimal conditions, can have recall lower than 60 percent due to variability in attention, fatigue, and interpretive subjectivity.[3] DLA Piper's PCaaS method can outperform full-scale manual efforts in efficiency, consistency, and cost-effectiveness. Our model also enables a tiered retrieval structure: It allows practitioners to calibrate review breadth (*e.g.,* by reviewing 1 or 10 percent of documents) against available resources while maintaining transparency about the tradeoffs between document volume and risk signal capture. In the compliance context, the goal of proactively monitoring unstructured data is to identify all areas of legal risk, rather than identifying the specific documents showing that risk. Our work suggests that such proactive monitoring is possible with 65 percent recall in 1 percent of documents.

Our PCaaS approach yielded actionable insights into areas of risk that may have otherwise gone unnoticed. Analysis of internal communications in real-world prospective datasets identified instances of potential non-compliance with internal procedures, as well as indications of inconsistent understanding among personnel regarding jurisdiction-specific rules under US state law.

These findings enabled the client to implement targeted interventions, including the development of revised training modules and adjustments to internal controls. Such enhancements

were designed not only to align workforce behavior with evolving compliance expectations but also to clarify areas of regulatory ambiguity that had contributed to inconsistent practices.

As a subjective measure of value, a stakeholder remarked that any of several identified "embers" could have grown into million-dollar reactive investigations and larger problems if left undetected and unaddressed. Importantly, the interventions and mitigations proposed by DLA Piper counsel in response to findings were generally low-touch and low-cost – tightening gaps in policies and education, among others. Because the issues were caught early, low-cost remedies were sufficient.

Subsequent data monitoring, conducted approximately 12 months after the initial review, indicated significant reductions in communications reflecting non-compliant patterns. At the same time, the continued deployment of PCaaS surfaced additional residual risks, allowing the client to adopt prospective mitigation strategies in real time. The result is a compliance posture that is not only more responsive but also increasingly anticipatory and preventative, grounded in continuous assessment of actual communications rather than reactionary episodic review or retrospective audit.

Industry recognition

Industry sources have praised DLA Piper's proactive compliance approach. RSGI noted general counsels "wanted the law firm to come to them and tell them what they didn't know about things they needed to know about," remarking DLA Piper was "taking [its] client's data in a safe, protected way, running Gen AI over it, and helping them know themselves better than they know themselves," and "the first law firm we got to know that was doing it at scale."[4] The *Financial Times* recognized DLA Piper's PCaaS work as 2024's Innovation in New Services to Manage Risk in its article, "*Fortune 10 Turns to DLA Piper for Legal Hybrid AI Solutions*." It noted that DLA Piper was "spearheading a trend that will see lawyers move away from advising clients based on precedents towards making more predictions about what may happen," and "thereby adds new and different meaning to knowing its clients, a top strategic aim."[5]

About DLA Piper

DLA Piper is a global law firm with 90 offices in over 40 countries, a *Chambers Global* Market Leader in Artificial Intelligence, and winner of *American Lawyer*'s 2024 Best Use of Generative AI and the *Financial Times*' 2024 Innovation in AI Strategy.

According to Chamber Global 2025, DLA Piper's global Artificial Intelligence and Data Analytics Co-Chair, Dr. Danny Tobey MD, JD, is "by far one of the top leaders and industry experts in AI." The *Financial Times* called him "a pioneer in the current shift in the practice of law from reactive to proactive."

Our US Artificial Intelligence and Data Analytics team includes more than 40 lawyers, data scientists, and software engineers, with more than 100 team members globally.

[1] John Carlin on stepping up DOJ corporate enforcement - Global Investigations Review

[2] A TAR experiment was conducted using active learning in a commercially available eDiscovery tool with a known violative seed set to find a known violative target set. Recall of 80 percent of the target set would require a review of the top 65 percent of scored documents, which is only marginally better than random.

[3] *See* TREC Legal Track (Text Retrieval Conference) results, where recall was often below 60 percent, especially in large, complex data sets, and Grossman & Cormack studies (2011 to 2014), which found that human reviewers frequently achieved recall below 60 percent.

[4] https://rsgi.co/app/uploads/2025/02/RSGI-Podcast-Episode-1-Transcript-1.pdf.

[5] Law firms lean into the business of prediction

dlapiper.com

Sept 08 2025 | MRKT0033894 PCaaS White Paper_v5LM