
Agentic AI Document Review Is Transformative for Complex Litigation

Released: March 21, 2025

Technical Contributors

Pei-Lun Tai <i>Syllo</i>	Jeffrey Chivers <i>Syllo</i>	Oz Ben-Ami <i>Syllo</i>
Jamie Callan ¹ <i>Language Technologies Institute Carnegie Mellon University</i>	Theodore Rostow <i>Syllo</i>	Amy Slagle <i>Syllo</i>
Nicolas Madan <i>Syllo</i>	Timothy Choi <i>Syllo</i>	Roy Liu <i>Syllo</i>
Frances Leggiere <i>Syllo</i>	Nick Caputo <i>Syllo</i>	Margot Feuerstein <i>Syllo</i>

Practitioner Contributors²

Jason A. Leckerman, Esq.
Thomas W. Hazlett, Esq.
Casey G. Watkins, Esq.
Ballard Spahr LLP

Brandon F. Renken, Esq.
Andrew C. Elkhoury, Esq.³
Anna V. Durham, Esq.
Mayer Brown LLP

Louis E. Dolan, Jr., Esq.
Vernon W. Johnson, III, Esq.
Brian A. Hill, Esq.
Michael P. Swiatocha
Anders D. van Marter
Anthony Vescova
Nixon Peabody LLP

Daniel S. Stromberg, Esq.⁴
Melissa Lardo Stewart, Esq.
Eliana J. Theodorou, Esq.
Outten & Golden LLP

David Stanton, Esq.
Rachelle L. Rennagel, Esq.
Pillsbury Winthrop Shaw Pittman LLP

Christopher D. Kercher, Esq.
Asher B. Griffin, Esq.
Heather Christenson, Esq.
Melissa A. Dalziel, Esq.⁵
Joanna D. Caytas, Esq.
Melissa Fu, Esq.
Paul Henderson, Esq.
Quinn Emanuel Urquhart & Sullivan, LLP

Joshua Upin, Esq.
Royer Cooper Cohen Braunfeld LLC

¹ Dr. Callan serves as a research advisor to Syllo.

² The Practitioner Contributors have contributed to this paper in their individual capacities.

³ Mr. Elkhoury's contributions occurred while he was an attorney associated with Mayer Brown, but he has since founded the law firm of Elkhoury Law PLLC.

⁴ Mr. Stromberg serves as a product advisor to Syllo.

⁵ Ms. Dalziel's contributions occurred while she was an attorney associated with Quinn Emanuel, though she has since taken a position with another firm.

Summary

The increasing volume of electronic documents in litigation has made document review one of the most significant drivers of cost and delay in modern legal proceedings. Previous methods to control costs and minimize delay—including outsourced managed review and non-generative technology-assisted review (TAR)—have limitations in granularity, accuracy, adaptability, and cost efficiency. Recent advancements in large language models (LLMs) have prompted eDiscovery professionals to begin using generative AI (GenAI) in the discovery process, and eDiscovery practitioners are becoming increasingly aware that LLMs are, at a minimum, a powerful tool in document review for investigations and litigations. However, straightforward applications of LLMs to large-scale, complex document reviews have encountered challenges due to limitations such as context windows, prompt complexity, terms of art, hallucinations, multi-document reasoning, limits on the number of issue codes applied, cost and time of iterative refinement, and overall expense. These problems are heightened in complex litigations involving large datasets, which has led some in the eDiscovery industry to conclude that LLMs have only a minor or supplemental role to play in large-scale, complex document review.

Syllo has developed an agentic AI system for document review that substantially overcomes these and other challenges. Syllo coordinates multiple LLMs that organize and delegate the work of the document review between one another and autonomously make decisions about how to conduct the review within guidelines set by users. This methodology delivers an automated document review solution that applies unlimited issue coding for large and complex document review projects in investigations and litigations. The agentic solution has consistently and substantially outperformed the benchmarks of prior generations of TAR in real-world complex litigations at a significant cost reduction compared to traditional managed review or managed review using prior generations of TAR. In the last ten completed responsiveness reviews by Syllo in live litigations, the lowest estimated Recall was 93.4%, the average estimated Recall was 97.8%, the median estimated Recall was 99.4%, and four of the reviews had estimated Recall of 100%. In the same reviews, the median estimated Precision was 85.9%, and the average estimated Precision was 79.7%.

When guided by sophisticated practitioners who have learned to use the system, agentic document review can provide a powerful strategic advantage in complex investigations and litigations, swifter and more accurate completion of document-review projects and deposition preparation, and a significant reduction in the overall cost of discovery in document-intensive cases.

I. Limitations of Prior Approaches to Complex Document Review

Over the last thirty years, as the use of electronic devices has proliferated, the document volumes in complex legal matters have routinely required the analysis of hundreds of thousands or millions of documents. As a result, document review has become the largest driver of cost and delay in modern litigation.⁶ The litigation industry adopted two principal strategies in response to this explosion of data: outsourcing managed reviews and non-generative TAR (*i.e.*, predictive coding) techniques. More recently, eDiscovery practitioners have begun to use generative AI to conduct linear review of documents. Each of these strategies has significant drawbacks.

A. Disadvantages of Outsourced Document Review

One response by large enterprises to the increasing cost of document review was to shift large-scale document review away from trial teams at outside law firms toward consulting firms or law firm subsidiaries, often referred to as Alternative Legal Service Providers (“ALSPs”). These ALSPs manage large teams of contract reviewers to perform first-level document review and document labeling at rates below those charged by law firms. The results of the outsourced first-level review are then passed back to the litigation team for second-level review, often with extensive (and expensive) back-and-forth cycles of quality control and cleanup workflows.

Three significant drawbacks of this outsourcing trend strike at the heart of sound litigation practices. First, outsourced review causes distributed knowledge of the factual record among numerous individuals who are not part of the trial team. Second, as the complexity and volume of the subject matter grows, the consistency and quality of the application of issue codes by review teams generally declines. Third, because it is overwhelming and time-consuming to manually review large and complex datasets for a large number of issue codes, the number of issue codes applied to datasets in managed review is generally limited. In addition, according to some studies, in complex cases, human review teams

⁶ John H. Beisner, *The Need for Effective Reform of the U.S. Civil Discovery Process*, 60 Duke L.J. 547 (2010), available at <https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1482&context=dlj>.

achieve average estimated Recall rates of roughly 60%, with upper-bound estimated Recall rates of roughly 80%.⁷ Recall measures completeness—the percentage of truly relevant documents successfully identified by the system out of all relevant documents in the dataset.⁸

B. Limitations of Non-Generative TAR (Predictive Coding)

Another approach to more cost-effective document review is non-generative TAR, which relies on non-generative machine learning techniques. In its early iterations (TAR 1.0), human reviewers label a “seed set” of the document universe, which the algorithm then uses to predict the labeling for other documents. Introduced in the 2000s, these approaches to TAR advanced through the 2010s. More recently, continuous active learning (CAL or TAR 2.0) has become more widely adopted. In CAL, the model is trained continuously (or at certain breakpoints) as reviewers code documents. This workflow eliminates the need for a seed set but still requires substantial review time to achieve acceptable results.

When properly used, TAR has provided substantial cost savings and quality improvement to the document review process. As litigants embraced these technologies, a body of case law emerged, setting forth standards for accuracy and validation. *See, e.g.*, The Sedona Conference TAR Case Law Primer, Second Edition (2023). Courts and litigants generally formed a consensus that 80 to 85 percent estimated Recall is an acceptable and legally defensible level of performance for predictive coding models in large cases, although the specific threshold in a given case also turned heavily on case-specific factors consistent with Federal Rule of Civil Procedure 26.⁹

Yet, non-generative TAR has significant limitations:

Coarser Document Analysis: In general, non-generative TAR models conduct a relatively coarse analysis of documents based on word frequency, concepts, and metadata features, as compared to modern LLMs, which evaluate the contextual meaning of phrases, sentences, and longer passages with a high degree of nuance. The ability of LLMs to handle these nuances gives them a significant advantage in document analysis over prior technologies.

Time-Consuming Startup: Non-generative TAR involves a substantial startup cost—*i.e.*, the human time required to label documents in a seed set or label documents to enable continuous active learning. In addition, predictive-coding models have required successive rounds of training, a time-consuming process. More modern predictive coding, such as CAL, continuously trains the model using reviewer coding but still requires substantial review time to achieve acceptable results. As noted in the Sedona Conference’s recent primer on TAR case law, litigants using TAR often run many iterations of review to achieve Recall rates in the realm of 70 to 80 percent.¹⁰

Limited Transparency: While non-generative TAR models generally attach a numerical score (*e.g.*, 1-100%) to a document’s likelihood of relevance, they do not provide an explanation as to why a document was suggested as relevant with that particular score. This lack of explanation requires second-level reviewers to start from scratch when confirming whether a particular document is, in fact, responsive and why. More systemically, if the predictive coding model has incorrectly tagged a series of documents, the lack of explanation makes it difficult to understand why the documents were miscoded and how to correct this miscoding.

Risk of Intercooder Disagreement: TAR’s reliance on human judgment may also lead to variability in how the model codes documents. Often, human reviewers will apply an issue code inconsistently, a phenomenon known as intercooder disagreement. Human biases can also come to bear, leading to fundamental shifts in relevance assessments as

⁷ Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, XVII RICH. J.L. & TECH. 11 at 37 (2011); Maura R. Grossman & Gordon V. Cormack, Technology Assisted Review in Electronic Discovery in Data Analysis in Law (Edward J. Waters ed. 2018). The contributors do not mean to disparage managed review teams. Reviewing large volumes of documents in complex litigation is simply an extremely hard thing for any group of people to do without the assistance of technology.

⁸ Practitioners also measure Precision—the percentage of documents identified by the system that are actually relevant.

⁹ Bolch Judicial Inst., Technology Assisted Review (TAR) Guidelines, January 2019, available at <https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1002&context=bolch>; *see also* The Sedona Conference TAR Case Law Primer, Second Edition (2023).

¹⁰ *See, e.g., ibid.*

the review proceeds. If a predictive coding model learns from such inconsistently coded documents, it too runs the risk of applying codes inconsistently or becoming confused by the mixed messages sent by disparate reviewers.

Limited Issue Codes and Adaptability: The need to train non-generative TAR models with human labeling makes it pragmatically difficult to apply a large number of issue codes to a complex dataset. Human review speed generally slows down as the number of issue codes increases, which places a practical constraint on the number of issue codes that can be applied with non-generative TAR. Similarly, the amount of time required to train and test predictive coding models is also a major drawback when new issues arise in the midst of the investigation or litigation. These limitations can restrict the usefulness of predictive coding models in more complex cases where the investigation objectives of the case team involve nuanced matrices of relevant facts, participants, categories, and timelines, which can evolve as new facts are uncovered as the discovery process unfolds.

C. *Shortcomings of Linear GenAI Document Review*

As LLMs have evolved, the legal industry has explored whether these models could improve upon the TAR predictive coding algorithms. Since modern LLMs are pre-trained on vast corpuses of data and can perform accurate document classification based on natural language instructions (*i.e.*, “prompts”), they require little to no training on case-specific documents or labeling by subject matter experts. The linguistic and conceptual nuance that is embodied in modern LLMs enables them to make distinctions and perform relevance predictions that significantly exceed TAR methodologies.

However, many of the efforts to utilize LLMs still involve a linear approach to reviewing and coding documents. For example, users provide an LLM with a single or multi-pronged prompt that sets forth case context and a description of what documents are responsive and/or the issues with which the document is to be coded. The LLM then considers the responsiveness of each document in the review population one by one.¹¹

This kind of linear GenAI review, while potentially adequate for smaller datasets (hundreds to thousands of documents), encounters significant limitations when applied to more complex and document-intensive matters:

Prompt Overload: A core challenge of linear LLM deployment in large-scale document reviews is the difficulty in creating a single, comprehensive prompt that accurately captures all relevant, nuanced factual issues, given the initial uncertainty inherent in complex cases. Splitting a complex prompt into multiple prompts (such as one prompt per issue code) can multiply the cost of review when each prompt is run linearly over the dataset. Further, the more complex and data-intensive the case is, the less certainty the prompt drafter has about the nuanced factual issues that may be hiding in that dataset, and the more complexity that needs to be packed into a prompt run linearly across the dataset. Attempting to address a multitude of issues in an aggregate prompt (such as a prompt comprising ten issue codes) increases the risk of overloading the LLM, diminishing its accuracy as it struggles to process all instructions simultaneously and creating a risk of inaccurate or incomplete coding. Improperly applied coding may require extensive quality control to detect, and necessitates additional, costly GenAI passes over the dataset to achieve a more accurate result. A complex, multi-faceted prompt requires more computational resources to review each document, resulting in a more expensive process overall.

Limited Issue Codes and High Cost for Large Datasets: The risks of prompt overload and cost overruns have caused some providers of GenAI document review software to place a limit on the number of issue codes or prompt length. However, when these limits are applied, a different problem arises. It is common in complex litigation for a party to be served with 20, 30, or more document requests, and when designing their own investigations, case teams often want to investigate dozens of issues or narrative threads. Document review approaches that are limited either technically or practically to a smaller number of issues thus do not meet the real demands of complex matters. Rather, they impose a constraint into which lawyers must artificially conform their investigation strategy to the limitations of linear GenAI review. This can lead users of linear GenAI review solutions to make compromises on the granularity of their review protocol. When more general issue codes are applied, human review teams must spend additional time sifting through these broad categories of documents to find relevant documents at the back end of the GenAI workflow. Broader issue codes could also require a re-run of an entire review or running multiple review passes to get closer to the more nuanced, desired results, which can rapidly increase the cost of the GenAI effort to the point of being cost-prohibitive.

¹¹ See, *e.g.*, Relativity & Redgrave Data, *Beyond the Bar: Generative AI as a Transformative Component of Legal Document Review* (2024).

Lack of Cost-Effective Adaptability: The reality of complex litigation is that strategic priorities and the perceived importance of given factual strains constantly shift and evolve as the case progresses. Linear GenAI does not readily adapt to this dynamism. A linear GenAI review process that requires subsequent “passes” across the dataset whenever a new issue arises is often cost-prohibitive as latency increases to intolerable rates. This fundamental disconnect from the innate dynamism of investigations and litigations significantly limits the effectiveness of linear GenAI approaches in complex cases.

II. Syлло’s Agentic AI Document Review: A New Paradigm

Syлло has created a novel, agentic AI system for document review that leverages an ensemble of varying-sized LLMs to conduct large-scale document analysis in complex investigations and litigations. Instead of a single-pass approach, Syлло orchestrates multiple LLMs performing distinct roles, including, among others, the roles of strategizing, determining next steps, performing quality control, extracting learnings from documents, synthesizing knowledge from learnings across documents, summarizing, and resolving inter-model disagreement. The behavior of the overall system and the dataflows involved in performing the document analysis are influenced by the outputs of the various LLMs working in concert.

The design and empirical studies of Syлло’s agentic approach indicate a higher ceiling on granularity, adaptability, context-sensitivity, cost-efficiency, and complexity as compared to prior non-GenAI and GenAI methodologies for large-scale document reviews.

The upshot for litigation teams is a document analysis system that can perform a highly accurate and categorized review to assist them at every stage of document analysis in litigation, including dataset culling, responsiveness review, subject matter issue coding, privilege review, and identification of hot documents. The benefits of this approach are numerous:

Dynamic Resource Allocation for Cost Efficiency and Accuracy: As the agentic review progresses, the telemetry of the system allows for observation of the amount of work performed by each LLM in each role. More complex and conceptually challenging datasets and documents will trigger more work performed by higher-end LLMs, whereas more straightforward review challenges will lean more heavily on the most cost-effective LLMs. This approach results in more efficient and less costly LLM application to complex document review and permits the ensemble of LLMs more freedom to determine which documents and parts of documents deserve a closer look for particular issues or nuances. It allows selective activation of more powerful or specialized models as needed to improve the quality of the review and minimizes reviewing completely irrelevant documents. This “division of labor” between LLMs mirrors the complexity of the document review project, akin to how complex reviews performed with outsourced reviewers resources might require more quality-control time and subsequent cleanup review hours.

Unlimited Issue Coding: Very significantly, an agentic approach cost-effectively accommodates an unlimited number of issue codes without causing prompt overload and without altering the system’s accuracy. This allows the granularity of the document analysis to match the number of requests for production or issues defined by the case team. As shown below, case teams routinely use Syлло to apply dozens of issue codes across large datasets. Also, for each label applied for each issue, the system provides a concise explanation of why a particular document (or its parts) is responsive to that issue with navigation to the relevant document content.

Swift and Cost-Effective Adaptability: Syлло’s agentic document review process does not require a seed set and does not require training a model each time the case team wants to add issue codes or change coding parameters. In fact, new legal issues can be integrated seamlessly, allowing on-the-fly adjustments without reprocessing the entire data set, simply by creating another natural language description of the issue to be investigated. The agentic system can leverage prior document analysis to accelerate and reduce the cost of these subsequent targeted queries. The result is a flexible and agile system that can adapt as the legal theories, facts, or other variables change in a matter. In addition, the work product created by the ensemble of models can be leveraged in subsequent analyses over the same dataset, and coding refinement (where the case team realizes an issue code was overly broad or unduly narrow) can be performed surgically at a small fraction of the cost of re-running an entire review.

III. Empirical Validation

Syлло’s AI systems have been used on active matters since 2023. In cooperation with law firm partners, the Syлло team has successfully completed more than 80 agentic document reviews in active litigation, with datasets spanning from thousands of documents to more than 2 million documents. Syлло has been used to apply dozens of issue codes in hours or days across hundreds of thousands of documents. These reviews have spanned numerous subject matters,

including antitrust litigation, environmental litigation, contract litigation, employment litigation, patent litigation, bankruptcy litigation, mass tort litigation, construction litigation, investment and shareholder disputes, M&A litigation, automotive litigation, real estate litigation, and insurance coverage litigation. Along the way, the Syлло team has developed standard workflows to leverage the capabilities of the agentic system. In the last ten completed responsiveness reviews by Syлло in active litigations, the lowest estimated Recall was 93.4%, the average estimated Recall was 97.8%, and four of the reviews had an estimated Recall of 100%. In the same reviews, the median estimated Precision was 85.9%, and the average estimated Precision was 79.7%.

A. *Syлло in Responsiveness Reviews*

1. Commercial Litigation

Ballard Spahr LLP (“Ballard”) represented an enterprise in a commercial litigation in which the Ballard client was interested in reducing the cost of its document review burden. Ballard attorneys educated the client on the option to use Syлло for the responsiveness review, and the client elected to move forward with Syлло as the solution for the responsiveness review.

Ballard’s client needed to respond to more than 30 requests for production served by the opposing party, and the collected dataset was more than 100,000 documents. With guidance from the Syлло team, the Ballard trial team articulated more than 25 issue codes that corresponded to the requests for production. It took approximately three hours of human time to set up the instructions for the review. No iteration was performed on the prompts based on human review of documents. Syлло’s agentic system applied more than 25 codes to the documents. More than 50,000 documents were identified by Syлло as responsive to one or more issues. The Ballard trial team performed precision testing and elusion testing and determined an estimated Precision of 95.56% and an estimated Recall of 99.4%.¹¹

Based on the performance of agentic review for responsiveness, the Ballard team also deployed agentic review on the opposing party’s production of more than 25,000 documents to identify deficiencies in the production. Syлло identified numerous specific gaps in the opposing party’s production, which enabled Ballard attorneys to demand a remedial production in a matter of days. The trial team also used Syлло to exclude non-responsive documents from the potential privilege documents and to make preliminary privilege calls to facilitate trial team review.

The Ballard team obtained a favorable resolution of the litigation for their client.

“We were able to complete a large document production with a high degree of confidence that we had identified all of the responsive documents,” said Casey Watkins, Of Counsel at Ballard. “Setting up the review was straightforward, and the review took a fraction of the time it would have taken if conducted using human review teams. Given the results we achieved and the amount the client paid for the total review, we were able to provide enormous value to our client.”

Ballard has subsequently employed Syлло in other document-intensive litigations, including for the application of more than 25 issue tags across a dataset of more than 1.5 million documents in a highly complex commercial litigation, which is currently the most complex litigation in which Syлло’s agentic review has been deployed.

2. Commercial Litigation

Joshua Upin, Esq., of Royer Cooper Cohen Braunfeld LLP (“RCCB”), was interested in performing a head-to-head comparison of Syлло’s agentic review capabilities against a managed review team leveraging CAL in an ongoing matter. The case selected was a complicated commercial litigation, involving hundreds of entities, many categories of commercial transactions, and more than 25 requests for production.

For the head-to-head comparison, both the managed review team and Syлло received the same review set of slightly less than 16,000 documents. This review set was randomly selected from the broader document population of more than 150,000 documents. The reviewers and the AI system performed their work using separate document coding platforms.

The complexity of the document review created challenges for the review team to accurately tag the documents. The review team required additional guidance and training from the trial team during the review, and based on second-level and quality-control reviews, there were multiple rounds of correction and re-tagging of the first-level review coding. Ultimately, the additional time and remedial work required for the outsourced review team to conduct their review of roughly 16,000 documents resulted in a per document review cost exceeding \$2.00 per document.

When the coding results were compared head-to-head, Syлло’s performance surpassed the human reviewers by a significant margin. The review team marked more than 5,400 documents in the head-to-head sample as responsive. Elusion testing against the review team revealed an estimated Recall rate below 67% and widespread miscoding that triggered several rounds of re-review. Syлло performed its review without any prompt refinement based on reviewing documents or overturns by the review team, and the RCCB team’s validation of Syлло’s coding revealed an estimated Recall of 93.44% and an estimated Precision of 69.81%.

“Our adoption of Syлло’s review solution has significantly reduced the time it takes for us to review our own documents and identify important documents in opposing parties’ production and provided better results than any other alternatives,” said Josh Upin, partner at RCCB. “In view of Syлло’s superior performance to human review and other available review platforms, it’s now my practice to use Syлло on document reviews of any significant size rather than hiring outsourced review teams. Learning and leveraging this technology enables me to get more quickly to documents that matter and better serve our clients while saving them significant expense.”

3. Commercial Litigation

A trial team at Quinn Emanuel Urquhart & Sullivan, LLP (“Quinn Emanuel”) was engaged as counsel in an accelerated litigation less than two months before trial was scheduled to occur. In the span of six weeks, the team needed to complete responsiveness reviews of more than 30,000 documents, review more than 40,000 documents produced by opposing parties, and complete depositions and pre-trial submissions.

For the responsiveness review, the Quinn Emanuel team initially defined more than 20 issue codes and prepared GenAI review instructions to use Syлло to perform the first-level document review. In a set of more than 30,000 documents, elusion testing confirmed that the estimated Recall was above 98% and the estimated Precision was above 74%. With respect to the review of opposing party productions, the trial team defined more than 40 issues for investigation and used Syлло agentic review to perform first-level review on a rolling basis. Finely categorizing the documents into 40 different categories allowed the team to find critical documents expeditiously, streamlining deposition and trial preparation. As the team identified new avenues of investigation, they defined new sets of review instructions, ran the instructions against the documents produced in the case, and were able to complete follow-up investigations within hours of identifying new issues.

Production in the litigation occurred on a rolling basis. Syлло’s ability to store previously defined issue codes and apply them to new productions and collections enabled the trial team to complete first-level reviews of new document sets within hours of their receipt. The trial team also used agentic review to identify deficiencies in the opposing party’s production. Given the timeframe of the litigation, this ability to rapidly identify gaps in productions and request supplementation ensured that the team had the evidence they needed to go to trial.

“Facing a high-stakes commercial dispute with only eight weeks until trial, our team needed to accomplish what seemed impossible—complete a substantial document review and production from our client, review opposing counsel’s documents to learn the case, and prepare for depositions on an extremely compressed timeline,” said Chris Kercher, a partner at Quinn Emanuel. “Syлло transformed our capabilities overnight. We rapidly identified and produced responsive materials from tens of thousands of documents to meet court-ordered deadlines, while simultaneously gaining unprecedented command over the adversary’s production. What truly differentiated Syлло was its ability to help us instantly adapt our review strategy as new issues emerged in the opponent’s documents, identify critical gaps in their production, and secure vital supplemental productions before deposition deadlines. In fast-moving, complex litigation where strategic advantage is measured in days, not months, Syлло transformed what would have been a logistics challenge into our strategic advantage.”

4. Employment Litigation

The plaintiffs’ employment firm Outten & Golden LLP has used Syлло to assist with many forms of document review and analysis. As one example, attorneys with Outten & Golden used Syлло to identify documents for production in a collection of 12,543 documents. The Outten & Golden team based their instructions to Syлло closely on the requests for production that had been served on their client in the case, resulting in the definition of 28 issue tags. One issue code was detected as overbroad as the system began its review, and that one issue code was re-drafted. Syлло applied the 28 issue tags across the documents and tagged 484 documents as responsive to one or more requests for production. An associate attorney with Outten & Golden conducted a second-level review of the documents tagged responsive and determined a Precision rate of 84.09%. The associate also performed elusion testing on the documents deemed non-responsive and found no documents in the null-set sample that were responsive, yielding an estimated Recall of 100%.

Based on numerous validations of Syлло’s document review solution, Outten & Golden uses Syлло’s agentic document review to review large client collections and certain document productions. The firm has observed results that exceed the standards for human and traditional TAR review.

“We’ve used Syлло’s automated document review function, and we were really impressed with the results,” said Melissa Lardo Stewart, Partner at Outten & Golden. “Syлло completed a review of thousands of documents in a few hours and our team’s review determined that it identified and labeled responsive documents accurately.”

B. Finding Hot Documents

In addition to reviewing documents for production, Syлло has been used to review the production record to identify hot documents for depositions, pre-trial motions, and trial. In these instances, the agentic reviews are conducted solely for the case team’s analysis. Case teams use Syлло’s agentic review system to identify decisive key documents from among vast swaths of responsive documents for a given issue.

1. Commercial Litigation

A trial team at Quinn Emanuel was engaged in a fast-paced litigation in advance of a preliminary injunction hearing in a commercial dispute. The team had less than a month to complete document productions and depositions. The trial team decided to use Syлло midway through its deposition preparation, as both a cross-check to make sure that key documents had been identified and to broaden their search to cover new, fast-arising issues more comprehensively than through keyword searching.

The team used Syлло to review more than 40,000 documents and to identify all documents that hit on more than 30 different issues. The trial team used Syлло for a zero-shot review (*i.e.*, the Syлло review ran once without the benefit of any prompt refinement based on document tagging performance). The Quinn Emanuel trial team performed precision testing and elusion testing on the zero-shot review, which confirmed that Syлло identified responsive documents with an estimated Recall of 98.69% and an estimated Precision of 92.83%. As to the issues that the Quinn Emanuel team had already reviewed, Syлло’s review confirmed the effectiveness of the trial team’s search of the document population. For the new issues, Syлло’s review identified more than 200 key documents that had not previously been identified.

“Syлло enables the streamlining of issues and organization of documents for complex litigation, allowing trial teams to move faster and more easily control the factual history of the case,” noted senior associate Paul Henderson. “Syлло’s agentic system reliably surfaces documents responsive to key issues and navigates substantial factual complexity better than any AI tool I have seen.”

2. Commercial Litigation

A trial team at Quinn Emanuel was faced with a tight timeline to prepare for depositions in a complex commercial litigation relating to the private equity industry. The team had already overseen an extended managed document review process and then the court allowed the opposing party to amend its pleadings months before the close of discovery. As a result, there were several new key issues that had not been the focus of a prior review.

The production universe in the case was more than 2 million documents. Syлло was first used to perform a targeted review to return a universe of the top 150 documents related to the new issues raised in the amended pleadings. The documents identified by Syлло were described by the trial team as “incredible.”

With depositions scheduled over the next four weeks, the Quinn Emanuel trial team next relied on Syлло to analyze the full production universe, including document productions that were produced after the review was complete, to provide a small number of “hot” documents responsive to 40 issue codes. Syлло’s agentic review system churned through the production universe to identify and narrowly return only the few hottest documents that related to any of more than 40 key factual propositions the trial team had identified in the leadup to depositions. Syлло identified, across six witnesses, 750 unique hot documents, 120 of which were newly identified in the case.

“The litigation situation we found ourselves in was familiar to many litigators—we had budgeted a certain amount for document review, and then the court’s decision changed the focus of the case,” said Melissa Dalziel (Of Counsel at Quinn Emanuel at the time and now Counsel at Alston & Bird). “Syлло allowed us to completely recalibrate our strategy and find a manageable number of the most relevant documents within a vast data set.”

3. Commercial Litigation

A trial team at Mayer Brown LLP was more than two years into a high-value litigation. The litigation involved nuanced issues of contracting and construction. More than \$300,000 had already been spent on managed document review in the litigation, and reviewers had already reviewed the production universe of more than 400,000 documents spanning more than 8 million pages.

Given the complexity of the document review and the stakes of the litigation, the Mayer Brown team wanted to ensure that key documents had not been overlooked as they prepared to enter a period of depositions. The Mayer Brown team articulated 15 primary issues to be addressed in depositions and trial, and they worked with the Syлло team to conform those issues into 15 issue codes for an automated first-level document review. Syлло completed the review, applying 15 issue codes across more than 400,000 documents in less than one week. Syлло’s agentic system identified slightly more than 18,000 documents as highly relevant to one or more of the issue codes and escalated a subset of hot documents for each of the issues.

Upon reviewing the hot documents identified by Syлло, the Mayer Brown team immediately identified critical documents that Syлло had escalated and that had not been previously identified in prior reviews. Ultimately, more than 50 documents that were not escalated by the managed review team but that were identified as highly relevant by Syлло were selected as hot documents by the trial team and slated for use in deposition, which represented almost 20% of the overall hot documents selected.

The Mayer Brown team concluded that having Syлло’s agentic review system perform a cross-check review in such a high-value litigation was more than worth the expense. “In high-stakes litigation, the prevailing party is often the one that is able to introduce the most compelling evidence to support their case,” said Brandon Renken, partner at Mayer Brown. “Apart from its speed and cost-effectiveness, Syлло more than proved its value by finding key documents that had been missed in the previously conducted managed review.”

C. Additional Applications in Litigation

Law firms have also successfully used Syлло’s agentic document review solution in other creative ways, such as complying with requirements to label production datasets by document request and to perform quality control analysis on human-reviewed datasets.

1. Labeling Every Document by Request for Production or Interrogatory

In a commercial litigation, Nixon Peabody LLP used Syлло to help satisfy a challenging directive from a tribunal to identify the request for production or interrogatory to which each of the 9,000 produced documents pertained. There were more than 30 requests and interrogatories, which translated into applying a coding palette of 30 issue codes. Prior to selecting Syлло for the project, Nixon Peabody began the review project with attorney and paralegal review staff, but due to the number of issue codes, the rate of review was not fast enough to meet the deadline. Nixon Peabody opted to use Syлло. At the time Syлло conducted the review, Nixon Peabody had just a few weeks to comply with the directive.

The document set was unique in that nearly every document was responsive (a richness of 100%). The Nixon Peabody trial team conducted a sample-based second-level review of the tagged documents to ensure that the tagging was correctly applied. Syлло expedited the process of review by providing rationales for the application of each tag. Nixon Peabody’s head of eDiscovery and litigation support, Mike Swiatocha, noted that providing the rationales for the tagging of each document inverts how documents are analyzed in the document review process, allowing second-level reviewers to focus on what is important in the case rather than labeling or summarizing documents. This allows for expedited review of the documents on a second-level review.

Not only was Nixon Peabody able to complete the second-level review in only a few days, but the trial team concluded that Syлло’s review had been highly accurate.

“Any attorney would have really struggled to complete the project for which we used Syлло, especially given the time pressures involved,” said Mike Swiatocha. “Syлло was the perfect solution because it could apply a superhuman number of issue codes to each document and apply them with impressive consistency.”

2. Early Case Assessment

A trial team at Quinn Emanuel was engaged in a bankruptcy proceeding and needed to perform early case assessment on multiple sets of document productions, totaling more than 20,000 documents, so they could advise their client on

the case posture and represent them at the proceeding. The review needed to be conducted on an expedited basis in advance of a hearing that was scheduled for a week after the team began ingesting documents onto the Syлло platform.

The Quinn Emanuel team deployed Syлло’s agentic system in two ways. First, they ran Syлло’s agentic document review over each population of documents as they were produced (more than 15 productions) from several parties to the bankruptcy proceeding. These reviews helped organize the documentary record around the key topics in the case, allowing fine-grained control over the documentary universe and helping the trial team quickly spot key issues that advanced their client’s interests, which was essential given the tight case deadlines.

Second, they identified more than 25 key facts that were central to their theory of the case and used Syлло’s agentic system to identify the 10 most relevant documents for each fact. This allowed the Quinn Emanuel team to quickly compile key documents for multiple depositions scheduled within days of receiving documents in rolling productions from multiple parties ahead of the hearing.

“Syлло is a groundbreaking platform that has quickly become my favorite document review tool,” said associate Joanna Caytas. “I was skeptical at the beginning, but Syлло delivered in a cost-efficient way what would have been very difficult to accomplish for a lean team on this timeline.”

3. Cross-Checking Reviewers When Responding to Interrogatories

Pillsbury Winthrop Shaw Pittman LLP (“Pillsbury”) utilized Syлло in a complex litigation to help analyze and identify documents responsive to contention interrogatories from a universe of more than 78,000 documents. This task was particularly challenging because the human review team had already used a set of 15 highly nuanced issue codes to categorize documents responsive to these interrogatories. Due to aggressive case deadlines that made it difficult to validate the review results in the time remaining, Pillsbury used Syлло to supplement the review team’s efforts in order to ensure completeness and accuracy.

Not only did Syлло validate the human review, but it also identified additional important documents that had been missed by human reviewers and highlighted certain documents the reviewers had coded inconsistently or incorrectly. Hence, the Syлло workflow provided an effective quality control mechanism that helped the team identify additional interrogatories to which a document might be responsive. In several instances, even though a human reviewer may have tagged a document as responsive to a single interrogatory issue, Syлло was able to identify additional issues relevant to that document. Importantly, these suggestions included an attribution that simplified validation by pointing out the specific pages or sections of each document that made it responsive to the issue code. Among the examples:

- One issue required fine-grained analysis and likely had few responsive documents. This was selected for a deep-dive evaluation. Reviewers coded 49 items responsive to the issue, and, of these, Syлло coded only seven responsive. The Pillsbury team reviewed the other 42 documents and determined that none of them were responsive to the issue.
- Syлло coded 10 documents as highly responsive to another issue, of which the human reviewers coded just one document as responsive. The Pillsbury team reviewed the ten documents and determined that all of them were probative of the issue.
- Syлло coded 73 documents as likely responsive to another issue, of which human reviewers had coded only six responsive. The Pillsbury team checked the other 67 documents and found most of them were, in fact, responsive to the issue.
- For another issue code, the reviewers coded 22 documents as responsive, but Syлло coded only 11 responsive. Pillsbury reviewed the 11 human-coded documents and determined that none of them were responsive to the issue in question.
- Syлло correctly identified 454 documents as very responsive to issues that had not been flagged by the review team (although they had been found responsive to other closely connected issues in the case), thereby demonstrating the platform’s ability to parse nuanced distinctions between related topics.

As a result, the Pillsbury team developed sufficient confidence in the system to begin to deploy it in more standard review workflows. “Syлло’s automated document review is reliable and provides unrivaled transparency into specific document characterization,” said David Stanton, a partner at Pillsbury. “Far from being a ‘black box,’ the tagging rationales applied by Syлло let us see why particular tags were applied to specific documents. This enabled workflows to adjust, optimize, and confirm the instructions we provided, and allowed us to very quickly leverage the insights we gained from its use.”

VI. Conclusion

Syllo's implementation of agentic document review has consistently and substantially outperformed the benchmarks of prior generations of TAR in real-world complex litigations at a significant cost reduction compared to traditional solutions. Syllo's implementation of agentic document review indicates superior granularity, adaptability, context-sensitivity, and complexity handling as compared to non-GenAI methodologies and linear GenAI methodologies. Case teams leverage Syllo's superior performance for responsiveness reviews at every stage of document analysis in litigation, including responsiveness review, subject matter issue coding, privilege review, and identification of hot documents. When guided by sophisticated litigators, agentic document review provides a powerful strategic advantage in complex investigations and litigations.