

# AI Privilege Log

*Apoorv Agarwal, Yi-Hsiang Kao, Antonio  
Limosani*

Text IQ

*Ali Shahidi, Lauren Doucette*

Sheppard Mullin Richter & Hampton LLP



# AI Privilege Log

## **Abstract**

The Sedona Conference Working Group on Electronic Document Retention and Production (WG1) 2020 Midyear Meeting will be held at The Camby Hotel in Phoenix, AZ, on April 30–May 1, 2020. This is a proposal for discussing a new AI based Privilege Log solution in the “Moving the Law Forward: Privilege Logs—Part II” Session.

## Explain Your Idea

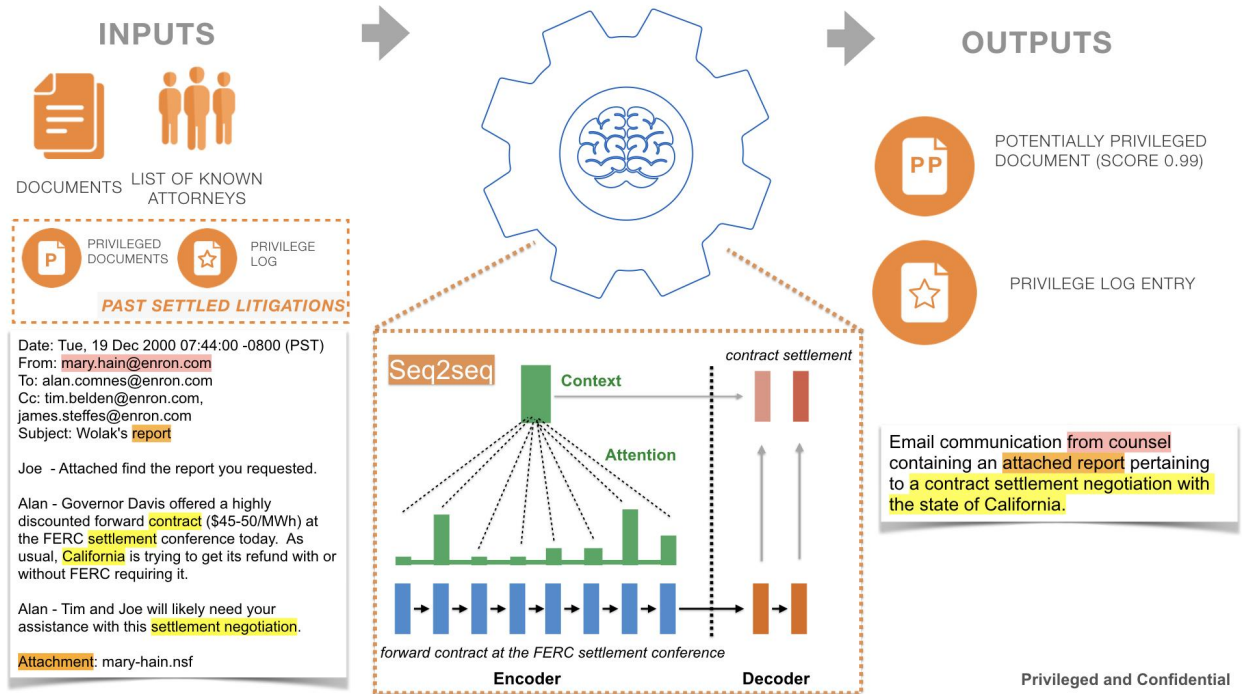
Our idea is to use a Sequence to Sequence (Seq2seq) AI-based system to produce narrative privilege log descriptions (or, “privilege log narratives”) akin to those produced by teams of contract attorneys. This system will drastically reduce, or potentially eliminate, the need for new narratives on live matters by leveraging the large number of available narratives present in past privilege log productions from settled matters. This idea is realizable with today’s AI technologies.

Seq2seq [1] represents a family of machine learning approaches used for language processing, where the specific component of Seq2seq to be used for automating the production of the privilege log narrative is known as “abstractive summarization of text.” This technology tool is capable of reducing much of the time and tediousness associated with privilege log production. It can be used to present and explain complex information to people who do not have the time, or other resources, required to analyze the raw data.

Information overload is a problem in modern digital society. It has been caused by the explosion of information produced on electronic communication channels. For textual information, this problem is even more significant due to the high level of cognitive engagement required for reading and understanding text. Automatic text abstractive summarization tools are thus useful in quickly understanding a large amount of information. Abstractive summarization requires deep understanding and reasoning over the text: determining the explicit or implicit meaning of each element, such as words, phrases, sentences and paragraphs; as well as then making inferences about their properties in order to, in turn, generate new text composing a succinct summary.

In this context, the construction of the privilege log narrative is dependent on abstractive summarization whereby email dialogue is translated into an informative and concise summary. Today, generic Seq2seq abstractive summarization models are available to perform this task. But, to achieve necessary high levels of accuracy and coherency given the informal nature of email dialogue, these models will need examples on which to adapt and fine tune their offerings. One could rely on a small coded set of privilege log narratives for emails on a given live matter, but, ideally, one would use data of past settled matters. Therefore, to build an advanced Seq2seq abstractive summarization model, past privilege log narrative data becomes extremely valuable.

A high level workflow of our Privilege Log idea is depicted in Figure 1, where data of both present and past matters is fed into the machine that has integrated within it Seq2seq algorithms enabling the generation of the Privilege Log.



**Figure 1: High level workflow of the generation of an AI Privilege Log.**

**How does your idea move the law forward in a reasoned and just way?**

Our mission is to help legal practitioners identify and protect sensitive information within large document collections. We have already developed a privilege review process, which uses machine learning to identify potentially privileged information within populations of unstructured data. This privilege review process helps clients significantly reduce risk -- by finding privileged documents that get missed by search terms (e.g. non-lawyers discussing legal advice given to them by an attorney, where no attorney is copied in the metadata) -- while also helping to reduce the overall time and cost of document production.

Our idea in this submission takes the privilege review process described above a step further beyond automated privilege identification to automated privilege logging. An automated privilege log tool facilitates a significant reduction in the human time necessary to execute this final step in document production. By maximizing the through-put of filling out the privilege log, while maintaining high fidelity and accuracy, our tool helps legal practitioners focus their time, effort, honed skill sets and intellectual abilities on the most "reasoned" tasks essential to the discovery process.

Given the explosion of data that legal practitioners must review on any particular matter, proven tools that can accurately address -- review and categorize -- sensitive information are becoming essential for legal practitioners. The solution discussed in this submission aims to move the practice of law forward by helping clients protect their privileged and sensitive data, while responding to discovery requests in accordance with their legal obligations under the Federal Rules of Civil Procedures as well as other discovery rules and obligations.

**How does your idea achieve the goals of enabling other parties to assess a claim of privilege or reducing the need for in camera examination of the documents?**

The Federal Rules of Civil Procedure, as well as other discovery rule regimes, require parties to identify withheld-for-privilege documents on a privilege log. The log must contain sufficient specificity such that the receiving party can readily understand the reasoning behind each claim of privilege. Under Federal Rule 26(b)(5), for example, parties are required to “describe the nature of the documents, communications, or tangible things not produced...without revealing information itself privileged or protected, [which] will enable other parties to assess the claim.”

Technology such as the automated privilege log tool described in this submission will help parties produce accurate and comprehensive privilege logs which are both responsive to discovery rule requirements, as well as informative towards parties evaluating privilege claims. Our automated privilege log narratives “reflect” the privileged documents identified through our AI machine-learning based privilege review process. Because our proposed privilege review process excels at comprehensively capturing the accurate universe of privileged documents in the data set, our automated privilege log thus reflects a comprehensive, accurate list of privileged documents accompanied by the requisite data fields (e.g., To; From; CC; BCC; Date Information; Subject Line) under discovery rules. Our proposed tool helps ensure the capture of a highly accurate universe of privilege documents, which, in turn, helps ensure highly accurate inputs into the privilege log.

Finally, our automated privilege log tool also reduces the time needed to produce the privilege log, which, in turn, helps facilitate a timely review of privilege claims among the parties.

**How does your idea align with Federal Rule of Civil Procedure 1's mandate to "secure the just, speedy, and inexpensive determination" of every case?**

Our idea helps legal practitioners identify sensitive data, particularly privileged documents, within a large unstructured data set, and speed up the process of explaining what makes these documents privileged. Protecting privileged information is a fundamental duty that legal counsel owe their clients. Technology that assists legal practitioners in fulfilling their obligation to protect privilege information facilitates "secur[ing] the just...determination" of their cases.

Our proposed automated privilege log tool aims to achieve faster production times, at significantly reduced costs, as compared to the time and cost associated with more traditional methods reliant on human reviewers. We have designed the process in a way that we can easily demonstrate, in objective terms, how the assistance of our privilege review technology will reduce risk, time and cost. The objectives of reducing risk, while also reducing time and cost associated with privilege review, are consistent with the Federal Rule of Civil Procedure 1's mandate to "secure the just, speedy, and inexpensive determination" of every case.

**Explain the practicality and technology implications of your idea.**

The process is transparent in so far that it leverages a limited amount of human review to maximum effect in applying that knowledge at scale to a full corpus of documents. At the very least much effort is reduced through auto-filling plausible and generally accurate privilege log descriptions. The reduced information allows for more judicious and practical filtering of potentially privileged documents.

Our idea allows for better understanding of the roles and functions of counsel with respect to the matter being litigated e.g. the log will detail the who, when and generally what was discussed in a minimal amount of time. This event meta-data allows for a basic narrative of privileged correspondences, relevant to the matter, to be gained, which can be more amenable to checks.

Employing automation over ten of thousands of documents is more practical than having each document individually read. Through this offering, we can help to define a future standard for "Technology assisted Privilege Log". The "AI Privilege Log" can become common-place and moreover spark innovation and continued improvements in this area. Given technology has facilitated the explosion of documents to review, it's only natural that technology be called upon to minimize the burdens and inefficiencies this has created.

## References

- [1] <https://en.wikipedia.org/wiki/Seq2seq>