

# Why Analytics Can Be Risky in the Wrong Hands

Having the most expensive or advanced tool in the toolbox doesn't matter if you don't know how to use it, and if you're not using those tools properly, there are risks everywhere.

BY BARRY SCHWARTZ  
FEBRUARY 2021

---

Recent years have seen dramatic advancements in e-discovery software, with some powerful analytics capabilities now often integrated as standard features instead of expensive add-ons. Simultaneously, e-discovery platforms have become increasingly user-friendly, with visually appealing interfaces that make it easier for users to view, manage and produce their data – or so they think.

In reality, the accessibility of analytics platforms can sometimes lead users to believe they understand the platforms better than they actually do. If a user is too emboldened, they may end up making a mistake that negatively impacts their case, such as not producing responsive documents or accidentally producing privileged information.

Having the most expensive or advanced tool in the toolbox doesn't matter if you don't know how to use it, and if you're not using those tools properly, there are risks everywhere. Here are a few examples of areas where inexperienced users tend to go wrong in their use of analytics and search tools:

- 1. Incorrect tagging of documents.** Analytics tools are only as good as the data that's fed into them, and a big mistake made by inexperienced users is to not tag documents correctly. A common example would be a responsive document that's attached to a nonresponsive email. It might seem correct to tag the email itself because of its attachment, as often was the normal practice before the use of analytics. However, because there's nothing responsive in the body of the email, tagging it will only make the analytics model less effective at best, or result in incorrect decisions by those tools at worst.
- 2. Over-reliance on basic search functions.** Search tools are powerful, but they're not foolproof. When running searches, inexperienced users sometimes overlook flaws in the underlying data that can affect the quality of their search results. Variations or misspellings of a name, or text that is contained in a screenshot or other image, can be easily missed.

*continued next page*



- 3. Incorrect handling of privilege review.** Sometimes attorneys clearly mark documents in a data set as privileged, but when they don't, a close read of a document may be necessary to ascertain that it's privileged – for instance, a brief mention in an email chain about the topic under discussion. Users shouldn't assume that an analytics model is going to automatically catch that. Some parties still do rely on clawback agreements that allow them to assert privilege over documents after they've been produced to the other party. But, as the old saying goes, you can't un-ring a bell. In some cases, if you try to claw back a document, the other party might claim that you've waived privilege by producing it, at which point you'd have a fight on your hands – a fight that is becoming harder to win. Recently, we worked with a client that was trying to produce documents on a tight deadline. They wanted to produce and send over everything that came up from a search in order to get the job done as quickly and as cheaply as possible. We advised the client that they were risking a waiver of privilege and successfully convinced the counsel to change course, thereby protecting the client's privileged documents. In general, if you have something you don't want the other side to see, don't rely on technology or **clawback agreements** alone to save you after you've already sent data out the door.
- 4. Incomplete production of documents.** Two standard analytics tools, **email threading** and deduplication, are often used to reduce redundancy in a data set before review. For instance, in an email thread that contains 20 messages, only the final email in the conversation would need to be included for review, since it fully includes all previous communications. However, depending on your production protocol agreement with the opposing party, you might need to produce all 20 messages in the conversation as separate documents, to make it easier to understand the flavor of a conversation. Not doing so could violate your production protocol order, and you would run the risk of sanctions from the court and/or increased costs from having to run the production over again to correct the deficiency. Producing only the most inclusive emails is not a bad strategy, but you must make sure that all parties have agreed to do so, and that they understand how to identify, handle and produce emails that branch out into separate discussion threads, too.

## The Math Behind the Tools

As straightforward as **analytics platforms** might appear on the surface, using them effectively comes down to understanding the math behind the tools. Risk and probability factor greatly into the work done by e-discovery professionals.

*continued next page*



## *Analytics Can Be Risky, continued*

Many leading analytics tools are built on a similar algorithm that actively learns from a representative sample of documents marked by the user either as responsive or non-responsive. The model built by those decisions is then applied to the remaining documents in the set, giving each one a score, typically from 0.1 to 0.9. (For instance, everything 0.5 and above might be considered responsive, and everything 0.49 and below would be non-responsive.) It's important to look below the non-responsive line and take a sample to see what percentage of those documents turn out to be responsive, and if necessary, further train the model. The goal is to achieve an acceptably low mathematical probability that there are responsive documents below the threshold.

This technical know-how works best when combined with an attorney review team that has a deep understanding of the case, the issues, and even institutional knowledge. Since analytical algorithms essentially magnify the decisions on a small subset of documents across the entire population, it's ever more critical that those base decisions are as accurate as possible. The more deeply the review team understands a case and the parties involved, the better their decisions on the base set will be, and the lower the probability of something important slipping through.

## **Every Case Is Different**

There's no one-size-fits-all in e-discovery; every case is unique and requires a customized approach. I've had decades of experience and so have my colleagues; yet, each of us have encountered situations that the others haven't. If one of us is presented with a legal matter that falls outside of our individual experiences, we'll talk it through as a team and debate the merits of different approaches as necessary to achieve the best overall result.

Using analytics tools without the proper know-how is like an amateur picking up a guitar and trying to play it. You might be able to strum a chord or two without a lesson, but getting the most out of your instrument requires knowledge and countless hours of practice. To get the best results out of your data, there's really no substitute for experience.



Barry Schwartz, Esq., CEDS, is senior vice president of advisory services at **BIA**, a leading national eDiscovery and digital forensics company. Schwartz is highly proficient in discovery and document review matters and holds more than 35 years of legal and business consulting management experience. He assists clients in a wide variety of areas including litigation and discovery, data retention and management, document review, regulatory compliance, privacy and cybersecurity. In addition to serving as BIA's general counsel and leading GDPR expert, Barry is a licensed attorney, a certified eDiscovery specialist (CEDS) and a certified Brainspace 6 analyst.

