

An Overview of Machine Learning Applications to Technology Assisted Review for E-Discovery Part Four: Ethical Considerations of TAR

[Sarah Bullard](#)

Consultant

This is the [fourth and final article](#) in a four-part series on technology assisted review (TAR), a process that uses machine learning to increase efficiency and decrease the cost of document review in discovery.

The first article discussed the differences between supervised and unsupervised machine learning algorithms, and why supervised learning algorithms are more commonly used with TAR for e-discovery. The second article explored two machine learning technologies pertinent to TAR: optical character recognition and natural language processing. The third article compared TAR with exhaustive manual review—that is, review performed solely by human beings with knowledge of the subject matter. This article considers two of the ethical concerns surrounding TAR: the importance of technical competence, and the disclosure of seeding sets.

Ethical Considerations of TAR

Technical Competence

In August 2012, the American Bar Association amended a comment on Rule 1.1 of the Model Rules of Professional Conduct, Competence, to include the following phrase: “To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, *including the benefits and risks associated with relevant technology...*” Since then, more than 30 states have adopted this amendment. This means that technological competence is

part of the ethical responsibility shared by all lawyers—and according to [an article](#) by attorney and legal commentator David Lat, this “relevant technology” includes artificial intelligence.

In 2015, the Honorable Joy Flowers Conti, Chief Judge of the U.S. District Court for the Western District of Pennsylvania, and Richard N. Lettieri of Lettieri Law Firm, wrote a helpful [guide to e-discovery ethics](#) in *The Federal Lawyer*, the magazine of the Federal Bar Association. This guide outlines “Nine Basic E-Discovery Skills” to help lawyers ethically navigate the e-discovery process. The first skill involves assessing the cost and necessity of e-discovery for a given case to ensure that the decision to use (or not to use) e-discovery will minimize costs. It naturally follows that a lawyer would need to consider whether TAR using machine learning technology would be the most cost-effective option for e-discovery, and, if so, which machine learning algorithm or product would be the most cost-effective option of those available. To answer these questions requires a working knowledge of current e-discovery technology.

Disclosure of Seed Sets

Artificially intelligent TAR must be used ethically and responsibly to prevent biased outcomes. Consider, for example, selection of seed sets by human beings. An unethical or unskilled human could provide a TAR system with a heavily biased seed set that results in the system ignoring certain types of documents which should have been identified as relevant. For this reason, there are many legal situations in which attorneys will be required to share their seed set with the court. This encourages accountability and transparency during the litigation process.

However, at times seed sets are not required to be disclosed. This is because some believe that seed sets are work product and should therefore be protected from discovery. This argument is reinforced by the fact that seed sets must contain both relevant and irrelevant documents from the corpus in question. Hence, disclosure of a seed set would mandate disclosure of irrelevant documents that would not normally need to be disclosed. This is discussed further in attorney [Shannon Kitzer’s article](#) in the first issue of the 2018 volume of the University of Illinois’ *Journal of Law, Technology & Policy*.

Regardless of whether disclosure of a particular seed set is required, it is ethically irresponsible to intentionally provide an e-discovery machine learning algorithm with a biased or incomplete seed set. This is implied in the Model Rules of Professional Conduct, which note that a lawyer

shall not “unlawfully obstruct another party’s access to evidence or unlawfully alter, destroy, or conceal a document or other material having potential evidentiary value. A lawyer shall not counsel or assist another person to do any such act...” Providing an algorithm with an intentionally biased seed set would likely constitute unlawfully concealing a document or other material having potential evidentiary value.

Although precision rates consistently drop as required recall increases, some methods of seeding are more effective for preserving precision than others.

It could also be considered ethically irresponsible to *unintentionally* use a suboptimal method to choose a seed set. The consequences of this latter scenario are discussed further in a [paper](#) by attorney and e-discovery expert Christian Mahoney et al. that compares and contrasts varying methods for choosing a seed set for a TAR algorithm. The paper explains that depending on the size of a seed set and the rate of recall required, different seeding methods will yield different precisions. Although precision rates consistently drop as required recall increases, some methods of seeding are more effective for preserving precision than others.

Additionally, larger seed sets consistently achieve better results than smaller seed sets. This may seem obvious, but it is still worth noting, because there is a motivator for those in charge of e-discovery to keep seed sets small: money. The costly time of experts is required to compile a seed set, and the larger the seed set, the greater the time needed to compile. However, Mahoney’s *Empirical Evaluations of Seed Set Selection Strategies for Predictive Coding* paper highlights an ethical responsibility not to cut corners on seed set development.

Conclusion

TAR can save time and money, but it requires that those involved be prepared to commit themselves to learning the new technology. Additionally, TAR is becoming more and more common in the world of e-discovery, and may one day be a *de facto* requirement of the process. Furthermore, more than 30 states now require that attorneys stay up-to-date with relevant areas of legal technology. Staying on top of issues involving artificial intelligence, TAR, and the creation and disclosure of seed sets is necessary for the modern legal practitioner to remain

ethical. For these reasons, it is a good idea to start growing familiar with TAR to ensure preparedness as it rises to ubiquity in the field of e-discovery.

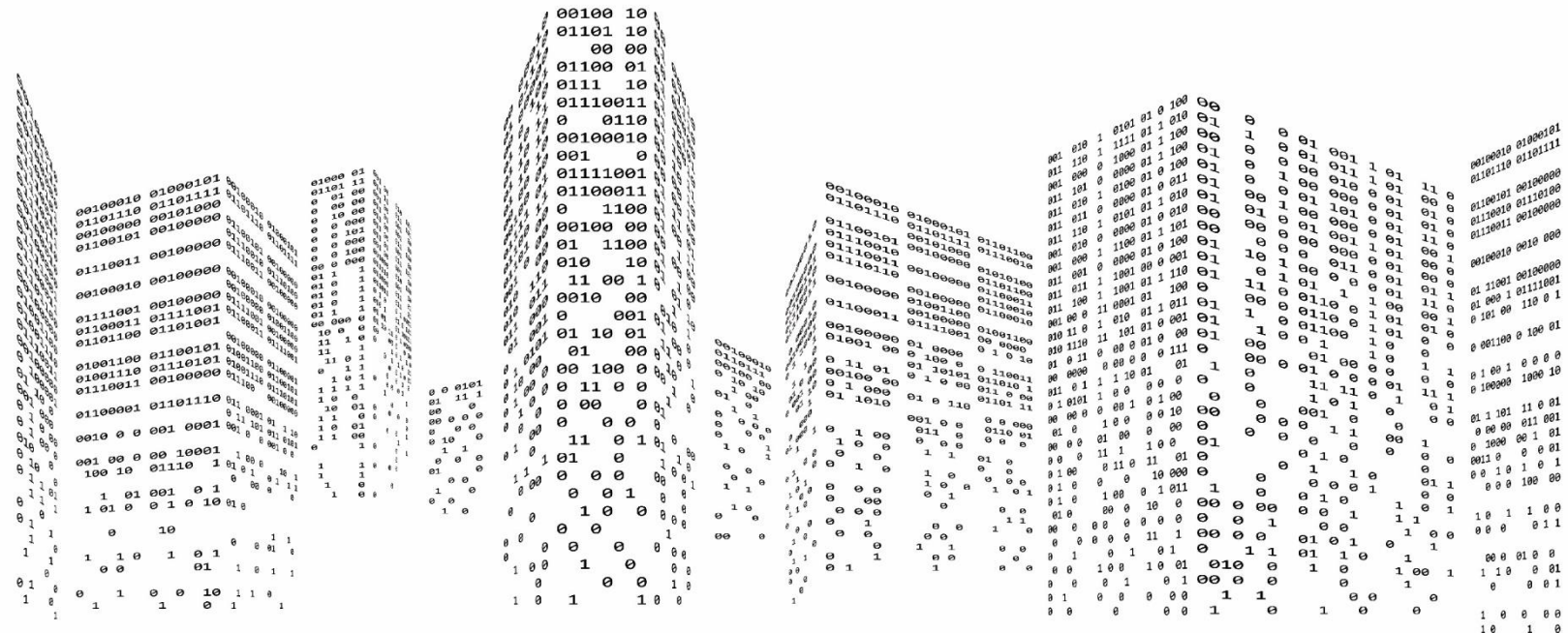
To learn more about DisputeSoft's e-discovery services including identification, recovery, preservation, and analysis of systems, databases, and other non-custodial evidence, visit our [electronic discovery services](#) page and explore a representative e-discovery case: [General Electric v. Mitsubishi Heavy Industries](#).

Sarah Bullard

Consultant



As a Consultant at DisputeSoft, Sarah assists in preparing expert, rebuttal, and investigative reports by reviewing disputed source code, databases, and computer systems, and by conducting analyses regarding source code quality and architecture. Her thorough understanding of object-oriented programming, data structures, and Unix systems allow her to effectively assist clients in the technical aspects of software-related matters, including matters involving allegations of software misappropriation.



If you are in need of an e-discovery expert, we invite you to consider [DisputeSoft](https://www.disputesoft.com).

Contact Information

Jeff Parmet, Managing Partner

301.251.6182

jparmet@disputesoft.com

12505 Park Potomac Ave. | Suite 475 | Potomac, MD | 20854