From: John Rabiej, J.D. ██████████████████
Sent: Thursday, December 19, 2019 10:19 AM
To: John Bates ████████████████████
Cc: 'Edward Cooper' ██████████████████; 'Marcus, Richard' ████████████████████; David Campbell ████████████████████; Robert Dow ████████████████████ >; Rebecca Womeldorf ██████████████████

Subject: Proposed Rule Amendment

Dear Chairman Bates:

The 2015 Rule 1 "cooperation" amendment sets out an ambiguous standard, which has caused recurring definitional disputes about the meaning of "cooperation" when applied to discovery. The ambiguity between "aspirational" and "obligatory" cooperation has resulted in parties avoiding TAR (technology-assisted review) rather than incur the costs of extended negotiations and satellite litigation resolving the ambiguity. Instead, the producing party falls back on less accurate and more expensive manual or other discovery methods.

TAR's paralysis will likely befall future machine-learning discovery tools because they entail the same functional steps and blackbox technology used in TAR. Most importantly, the parties' abstention from TAR has thwarted the development of a body law on the meaning and limits of "cooperation." The issue is rarely formally presented, and courts are denied the opportunity to develop common-law guidance.

I propose that the committee undertake a study with a view to amending Rule 16 or Rule 26 to provide greater clarity on the extent of party cooperation in discovery.

Rationale for Rule 1 "Cooperation"

The extensive committee records, including agenda books and meeting and mini-conference minutes, show that significant concerns were raised about amending Rule 1 to promote "cooperation." Some expressed concern about the committee's inability to define precisely what "cooperation" in discovery actually means. Others noted that any amendment promoting cooperation would "raise a vast array of questions that may be inconsistent with the adversary system of justice." They feared that it would be "difficult to identify a proper balance of cooperation with legitimate, even essential, adversary behavior." Others passionately believed that cooperation made common sense and was critical to promote reasonable, proportional discovery. The committee itself had mixed feelings about enforcing "cooperation" by rule.

In the face of the polar divide in the bar, the Committee deliberately placed the reference to "cooperation" in the Committee Note, instead of the rule text, as many had recommended. By so doing, the Committee intended to promote "cooperation" as an aspiration, while sidestepping directly imposing a "cooperation" duty by rule text. The committee expected that the rule amendments would encourage the bench and bar to develop case law and practice, which would evolve and provide clearer guidance in time.

Competing Positions

A party producing discovery using TAR typically discloses the following information: (1) name of TAR software; (2) data sources; (3) date-range restrictions; (4) deduplication methods; (5) excluded files; (6) number of documents subjected to TAR; (7) estimated richness of the TAR set; (8) number of documents classified as responsive; (9) number of documents classified as non-responsive; and (10) recall rate achieved.

In addition to these disclosures, the adversary party requests additional cooperation, including some control of the administration of TAR, which has stymied progress, regarding: (1) culling and narrowing the universe of documents to be searched; (2) training the algorithm-based blackbox; and (3) verifying the validation of the algorithm blackbox's results.

One side objects, contending that no Federal Rule of Civil Procedure, including aspirational Rule 1, imposes a "cooperation" duty. They contend that parties continue to be subject only to the same rules that traditionally apply to paper discovery, which impose no cooperation duty and permit the producing party to use whatever discovery means they believe are adequate, subject to court review *only if* errors are shown. Under this position, a producing party has no obligation to disclose to the other party how it actually conducts its discovery, including how it trains and validates TAR. It contends that the traditional Rule 26(g) duty and the risk of failing to provide relevant discovery shown on a motion to compel, subjecting the party to possible court sanctions, provide sufficient checks on the producing party.

The other side contends that the committee note ties cooperation into the Rule 1 amendment, obligating parties to apply the rules to secure speedy and inexpensive results. Under this position, the need for cooperation when using machine-learning, black-box technology discovery is especially warranted. Lawyers generally exercise significant judgment when they apply TAR, which is subject to self-interest and bias. They contend that under these circumstances both parties must have the opportunity to be involved in the actual training and validation TAR processes to ensure fair results.

 Specific Showing of "Cooperation" Definitional Disputes When Applying TAR

For more than 12 months, a volunteer team of 20-30 defendant and plaintiff lawyers and judges has worked unsuccessfully to develop a TAR protocol under the Duke Bolch Judicial Institute's auspices. We have failed to reach consensus on the extent of cooperation regarding three essential TAR steps, which portend similar stalemates, because all future machine-learning AI-driven discovery tools will entail these same three steps.

First Functional Step – A responding party typically informs the other party of the keywords it will use to cull documents in discovery. Often, the requesting party goes a step further and demands that certain keywords be added or modified. The request often expands the Boolean-based search by substituting a suffix with "***," which exponentially increases the volume of hits.

Second Functional Step -- A human trains TAR by classifying a series of documents responsive (relevant) or non-responsive (non-relevant), by using software and neutral algorithms available in the general marketplace. The algorithms detect patterns in the documents classified by the human as responsive and TAR begins independently to classify responsive documents, which are subject to review and correction by a human. TAR's classifications become more accurate as the training process proceeds.

Although algorithms used by TAR software, which are available in the open marketplace, are generally neutral, the training of the system and the classification of individual documents require human judgment, which is subject to inadvertent or deliberate bias.

After the responding party informs the other party about how it will train the TAR, the requesting party side goes a step further and demands that it actually play a part in the training. Under TAR 1.0, the requesting party demands to add or modify the documents used in training TAR. Under TAR 2.0, the

requesting party demands an opportunity to review and modify the criteria used by the producing party when determining whether a document is responsive.

Third Functional Step -- After informing the other side about how the TAR results will be validated generally, the requesting party goes a step further and demands to examine the random sample of documents used as part of the final quality control (null set), which TAR categorized as non-responsive.   They contend that the producing party must disclose not only the responsive documents that TAR incorrectly omitted in the random sample, but also the documents that TAR categorized as non-responsive in the null set. The requesting party claims that only by also disclosing the non-responsive documents can it be assured that the backbox TAR results have been properly verified.

Under Rule 1, one side asserts no obligation to accommodate, while the other side asserts judges are requiring parties to compromise, regarding all three functional steps.

Stalemate in Case Law and Practice

Each side is armed with arguments based on Rule 1 to support its position. These arguments come to the fore when a producing party faces demands that it not only must consider, but also must modify, its use of TAR to take into account the other side's recommendations regarding: (1) the keywords used to cull the documents; (2) the documents to be used during training in TAR 1.0 or the criteria to be applied in TAR 2.0 in determining responsiveness; and (3) disclosure of the non-responsive (non-relevant) documents identified in the null set as part of the quality control.

The case law and practices that the committee expected to develop from court decisions resolving these respective positions has not, and will not, happen.   Rather than risk the costs and burden of extended negotiations with counsel and satellite litigation, which may lead to repeating the burdensome and costly TAR process, producing parties avoid using TAR and opt instead to rely on manual or other discovery methods.

The same fears and the same stalemate will result when future machine-learning AI-driven technology discovery tools emerge, which entail the same three functional steps, i.e., culling, training, and validation/verification. The upshot is that the courts will continue to be unable to develop a common-law solution.

Mass-Tort MDL TAR Protocol Exemplars

Parties have agreed to TAR protocols in several high-profile, mass-tort mega MDLs, which provide the requesting party substantial rights in actively participating in and control of the producing party's culling of documents and training and verifying of the validation TAR process. One side recommends universal adoption of the protocol as ideal in all cases, while the other side distinguishes such protocols because in such large cases TAR is a practical necessity and it is subject to intense judicial case-management. They assert that the example is inapposite because the safeguards inherent in such mega mass-tort litigation are not present in other more routine litigation.

Solution

A carefully crafted amendment to either Rule 16 or Rule 26 can calibrate the party cooperation required in discovery to the extent the committee determines optimal. Such calibration would need to account for the extent of court involvement in resolving disputes that may be raised at every stage that the amendment requires party cooperation.

At one end of the spectrum, an amendment could require only that parties consult with each other about discovery, disclose discovery tools and techniques, and allow suggested modifications, but not force the producing party to accommodate any of the suggestions. In the TAR scenario, a party could suggest modifications of the keywords used for culling, the disclosure of documents or criteria used in TAR

training, and the disclosure of documents classified as non-responsive in the null set. The producing party could consider the suggestions, but would not be obligated to accept any of them. A court could would likely take into account the producing party's rejection of all suggestions if the party is found ultimately to have omitted production of responsive relevant materials.

At the spectrum's other end, a rule amendment could require the parties to consult with each other, agree upon discovery tools and techniques, and obligate the producing party to accommodate suggestions from the other party by making appropriate changes in their discovery, unless the dispute is raised to the court for resolution. In the TAR scenario, a party could suggest modifications of the keywords used for culling, the documents or criteria used in TAR training, and the disclosure of documents classified as non-responsive in the null set. The producing party would be obligated to accommodate the suggestions, unless they raise the matter with the court. Any failure to provide responsive relevant materials in the end would more likely be excused.

Conclusion

The committee can adopt a rule amendment that represents either end of the proposed cooperation spectrum or something in between. Whatever the committee decides, it is critical that the guidance on the extent of party "cooperation" is clear when applied in actual discovery. And in the case of TAR, the guidance should leave no question about the producing party's respective duties in operating, training, and validating TAR.

The effects of Rule 1's ambiguous meaning of "cooperation" on TAR is but a harbinger of what we can expect regarding future AI-driven, machine-learning discovery tools. Unless more concrete guidance is provided on what "cooperation" actually means, the problems that we have encountered in developing a TAR protocol will continue to arise and become more pronounced when newer AI-driven, machine-learning discovery tools are brought to the market. The stalement will persist for no good reason, and efficiency and cost savings will continue to be lost.

Thank you for considering my proposal, and Merry Christmas and Happy New Year!

John K. Rabiej