

11-2023

Disclosing Machine Inputs and Outputs: The Vulnerability of Legal Technology in Civil Discovery

Joshua Concannon

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/njtip>

Recommended Citation

Joshua Concannon, *Disclosing Machine Inputs and Outputs: The Vulnerability of Legal Technology in Civil Discovery*, 21 NW. J. TECH. & INTELL. PROP. (2023).

<https://scholarlycommons.law.northwestern.edu/njtip/vol21/iss1/4>

This Article is brought to you for free and open access by Northwestern Pritzker School of Law Scholarly Commons. It has been accepted for inclusion in Northwestern Journal of Technology and Intellectual Property by an authorized editor of Northwestern Pritzker School of Law Scholarly Commons.

N O R T H W E S T E R N
JOURNAL OF TECHNOLOGY
AND
INTELLECTUAL PROPERTY

**DISCLOSING MACHINE INPUTS AND
OUTPUTS: THE VULNERABILITY OF
LEGAL TECHNOLOGY IN CIVIL
DISCOVERY**

Joshua B. Concannon



DISCLOSING MACHINE INPUTS AND OUTPUTS: THE VULNERABILITY OF LEGAL TECHNOLOGY IN CIVIL DISCOVERY

*Joshua B. Concannon**

ABSTRACT—Smart technology has begun to infiltrate nearly every corner of society. While the legal profession managed to resist this intrusion relative to other industries for many years, it is now undeniable that machines frequently supplement lawyers and civil procedures such as discovery will need to adapt. As litigants, usually the well-resourced ones, increasingly utilize machine intelligence, concerns about accuracy and unfair advantage have sprung up on the other side of technology use. Information asymmetry is exacerbated when technology is accessible to only one party, and, consequently, curious litigants may seek discovery about the technology’s implementation in the context of the dispute. Thus, as law firms and corporate legal departments consider whether and how to integrate emerging technologies into their operations, it will be important to know their exposure to litigation. This paper provides suggestions for whether, and to what extent, parties should be able to obtain discovery about an opponent’s tools, including their machine inputs and outputs.

After reviewing the discovery process broadly, this paper will walk through three of the most common and relevant technologies, including technology-assisted review in discovery productions, predictive analytics for forecasting outcomes, and document generation for creating legal documents. If a party demands to see the details behind the use of one of these technologies, should the court compel production? This paper offers several considerations for judges in exercising their discretion to answer this question.

I. INTRODUCTION	130
II. PURPOSE OF DISCOVERY	132
A. <i>The Broad Goal of Discovery</i>	132
B. <i>Relevance and Proportionality</i>	133
III. PROTECTIONS FROM DISCOVERY	135
IV. APPLYING PRINCIPLES OF DISCOVERY TO EMERGING TECHNOLOGY	138
A. <i>Technology Assisted Review (“TAR”)</i>	139

* Northwestern Pritzker School of Law, J.D., 2024.

B. Predictive Analytics	145
C. Document Generation.....	151
V. CONCLUSION	153

I. INTRODUCTION

These days, we see smart technology integrated into nearly every corner of society. Companies like Netflix and Amazon use predictive analytic software to make recommendations that fit our preferences,¹ and search engines like Google fill in our search terms automatically. Similarly, Facebook and other social media platforms use machine learning to personalize our news feeds and ad exposure.² But now imagine that instead of technology recommending the next show to watch or article to read, it is recommending the forum in which to file your lawsuit, whether to settle a case or go to trial, or which data should be extracted for review to satisfy an adversary’s discovery production request. Technologies like these already exist and they are increasingly being used to prepare for litigation or are the subject of litigation. While it may not be the case that lawyers and judges will soon be replaced by robots, their work is increasingly supplemented by technology and civil procedures, like discovery, will need to adapt.³

As technology has started to augment litigants, doubts about its performance and accuracy, as well as concerns about unfair advantage, have sprung up between adversarial parties. When technology is accessible to only one party, information asymmetry is exacerbated. Consequently, litigants become curious about the technology’s implementation in the context of a dispute and might seek discovery on it. As law firms and corporate legal departments consider whether and how to incorporate emerging technologies into their operations, it will be important to know how exposed they may be in litigation. For example, discovery of technology tools may present privacy concerns. Additionally, it could diminish the technology’s potential

¹ Richard Finkelman & Karl Schliep, *Anticipating Predictive Analytics’ Potential Uses in Litigation*, LAW360 (Feb. 09, 2021), <https://www.law360.com/articles/1353409/anticipating-predictive-analytics-potential-uses-in-litigation> [<https://perma.cc/AJ2B-LT2N>].

² Bernard Marr, *The 10 Best Examples of How AI Is Already Used In Our Everyday Life*, FORBES (Dec. 16, 2019), <https://www.forbes.com/sites/bernardmarr/2019/12/16/the-10-best-examples-of-how-ai-is-already-used-in-our-everyday-life/?sh=6ec477771171> [<https://perma.cc/CNX8-UVZ5>].

³ See David F. Engstrom & Jonah B. Gelbach, *Legal Tech, Civil Procedure, and the Future of Adversarialism*, 169 U. PA. L. REV. 1001, 1004-05 (2021) (explaining the need for the civil procedure and inner workings of the adversarial system to keep up with the legal profession as it adopts new technological developments).

competitive benefit to the user. Thus, practitioners may be less compelled to invest in a technology if they must share everything with opposing counsel.

This paper explores the question of how courts and litigants ought to think about the civil discovery process in light of the new realities of legal technology. It will provide suggestions for whether, and to what extent, parties should be able to obtain discovery about an opponent's tools, including their inputs and outputs. This paper will focus specifically on the technologies that are most used in litigation or are likely to be subjected to a discovery request, which include technology assisted review ("TAR") in e-discovery, predictive analytics, and document generation. TAR, which uses technology to review large sets of data and documents for production for discovery, is replacing or substantially supplementing junior associates who traditionally did these reviews manually.⁴ We have already seen such technology assisted discovery production methods be subject to their own discovery, so-called "discovery on discovery."⁵ While TAR is one of the most widespread uses of legal tech, computational services have also begun to venture into other tasks, such as predicting outcomes and generating documents. Outcome prediction tools use data to guide an attorney in his or her decision making, whether it be with selecting the ideal forum or identifying the most favorable settlement terms. In civil discovery, a party might request that its adversary produce the data analytic details it has used leading up to the litigation. Document generation tools assemble documents of various sorts, from contracts to legal briefs. As these become more common, they will likely become the subject of various types of discovery requests too.

Before diving into the technologies themselves, this paper first discusses the discovery process broadly to lay out the framework in which the discoverability of the tools will be addressed. The paper then narrows in on discovery protections, including the attorney-client privilege and the work-product doctrine. Our focus will be on the work-product doctrine, which is the presumptive defense to many of these requests. The next part discusses the first relevant legal tech, TAR. We will explore what TAR is and how it functions in the pre-trial process before suggesting a lens through which the question of its discoverability should be viewed. The last two parts will do the same analysis with outcome prediction and document generation.

For the reasons discussed in this paper, judges should exercise their discretionary power to decide whether to compel discovery upon a motion

⁴ John O. McGinnis & Russel G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services*, 82 *FORDHAM L. REV.* 3041, 3041 (2014).

⁵ See *infra* notes 71-76 and accompanying text.

from one of the parties in light of several considerations, including (1) the subject of the production and whether it is relevant to the underlying claims in the dispute or to assessing the adequacy of the adversary's production; (2) the type of attorney-work that is at issue and whether it's the type that risks exposing the attorney's mental impressions; and (3) the lifecycle of the development and usage of the technology and whether the attorney judgment involved was exercised in anticipation of litigation.

II. PURPOSE OF DISCOVERY

A. *The Broad Goal of Discovery*

Discovery is widely recognized as the foundation of American litigation.⁶ It is the pre-trial phase when all material facts necessary to prove or disprove a claim are identified.⁷ Discovery happens at the outset of a lawsuit and serves as the backdrop for everything that follows, so cases are often won or lost based on what happens in this stage.⁸ In elaborating on the best way to treat discovery requests on legal technologies in the future, it is important to identify the normative goals that the rules of discovery are designed to achieve. Describing these goals provides the measure by which to gauge the success of the emerging approaches.⁹

The general purpose of discovery is to drive the disclosure of all relevant information so that the ultimate resolution of disputed issues is based on a full and accurate understanding of the true facts, and therefore embody a fair and just result.¹⁰ Significant discretion is traditionally given to trial court judges to administer the discovery process in order to achieve this goal.¹¹ The court must "tailor discovery to the circumstances of the case at hand, to adjust the timing of discovery, and apportion costs and burdens in a way that is fair and reasonable."¹²

⁶ Diego A. Zambrano, *Discovery as Regulation*, 119 MICH. L. REV. 71, 72 (2020).

⁷ Engstrom, *supra* note 3, at 1043.

⁸ *Id.* For these reasons, discovery is a central focus for many litigators and afforded significant time and resources.

⁹ Seth K. Endo, *Technological Opacity & Procedural Justice*, 59 B.C. L. REV. 821, 827-828 (2018).

¹⁰ TIC Park Ctr. 9, LLC v. Cabot, No. 16-24569-CIV, 2017 WL 3034547, at *2 (S.D. Fla. July 18, 2017); Peterson v. Wright Med. Tech., Inc., No. 11-1330, 2013 WL 655527, at *6 (C.D. Ill. Feb. 21, 2013) ("The very purpose of discovery is to give the parties the opportunity to learn what their opponents know about the issues in the case.").

¹¹ *See, e.g., Gov't of Ghana v. ProEnergy Servs., LLC*, 677 F.3d 340, 345 (8th Cir. 2012) ("[I]n the discovery arena the trial judge's discretion is particularly broad").

¹² Witt v. GC Servs. Ltd. P'Ship, 307 F.R.D. 554, 559 (D. Colo. 2014).

All discovery requests will fall into either the category of “merits-directed discovery” or of “process-directed discovery.”¹³ The former is directed to the substance of the litigation and is framed by specific elements of the claims; it focuses on issues relevant to case resolution.¹⁴ The latter is directed to the production process itself and its efficacy.¹⁵ Different types of legal technologies implicate each of these types of discovery. For example, discovery related to the selection criteria used to gather electronically stored information (“ESI”) is a process-directed inquiry. However, discovery of the inputs used in generating a contract at the heart of the litigation is a merit-directed inquiry.

B. Relevance and Proportionality

While information symmetry is the aim of discovery, it has a counteracting goal of expedience and cost-efficiency.¹⁶ Maximal exchange of information has its benefits, but time and cost concerns remain, hence why the Federal Rules have relevance and proportionality thresholds.¹⁷

These priorities largely stem from exorbitant discovery costs, which remain a substantial source of litigation expenses. Some estimate that they make up one-quarter to one-third of the total costs a party incurs to litigate.¹⁸ Importantly, parties typically bear their own costs in American litigation.¹⁹ This enables parties to externalize costs onto adversaries by requesting more information than necessary and forcing them to bear the cost of producing it. Similarly, a litigant can bury relevant facts in a mountain of information it produces and force the other party to bear the cost of sorting through it.²⁰

The concern for cost-efficiency is embodied in the Federal Rules generally, but especially in their relevancy and proportionality requirements for discovery productions. Federal Rule 1, the foundation of the civil procedure rules,²¹ states that the rules are to “be construed, administered, and employed by the court and the parties to secure the *just, speedy, and*

¹³ For more information on the differences between these two types of discovery requests, see Craig B. Shaffer, *Deconstructing “Discovery About Discovery,”* 19 SEDONA CONF. J. 215, 217 (2018).

¹⁴ *Witt*, 307 F.R.D. at 559.

¹⁵ *Id.*

¹⁶ *See, e.g.,* Fields v. Roberts, No. 1:06-cv-00407-AWI-GSA-PC, 2013 WL 6053824, at *1 (E.D. Cal. Sept. 16, 2013).

¹⁷ FED. R. CIV. P. 26(b)(1).

¹⁸ Engstrom & Gelbach, *supra* note 3, at 1048.

¹⁹ Alyeska Pipeline Serv. Co. v. Wilderness Soc’y, 421 U.S. 240, 257 (1975) (“[A]bsent statute or enforceable contract, litigants pay their own attorneys’ fees”).

²⁰ Engstrom, *supra* note 3, at 1047.

²¹ Shaffer, *supra* note 13, at 225.

inexpensive determination of every action and proceeding.”²² To strike a balance between the “just,” “speedy,” and “inexpensive” ideals, the rules focus on the relevance and proportionality of discovery requests.²³ Rule 26(b)(1) provides the relevance and proportionality limits. It states, “[p]arties may obtain discovery regarding any nonprivileged matter that is *relevant* to any party’s claim or defense and *proportional* to the needs of the case.”²⁴

The non-moving party has the burden to show a lack of relevance by demonstrating that the request either does not come within Rule 26(b)(1)’s broad scope of relevance or is of such marginal relevance that the costs of discovery outweigh the ordinary presumption that favors broad disclosure.²⁵ Relevant information need not be admissible at trial if the request appears reasonably calculated to lead to the discovery of admissible evidence.²⁶ However, the standard under Rule 26(b)(1), while broader than the admissibility standard under Rule 401 of the Federal Rules of Evidence,²⁷ does not permit a requesting party to engage in a fishing expedition hoping to turn up something useful.²⁸

Rule 26 was amended in 2000 to acknowledge that “[a] variety of types of information not directly pertinent to the incident in suit could be relevant to the claims or defenses raised in a given action.”²⁹ This includes “[i]nformation about organizational arrangements or filing systems of a party . . . if likely to yield or lead to the discovery of admissible information.”³⁰ Thus, when the courts think about relevant information in the context of discovery, they do not limit themselves to substantive facts.³¹

²² FED. R. CIV. P. 1 (emphasis added).

²³ FED. R. CIV. P. 26(b)(1).

²⁴ *Id.* (emphasis added).

²⁵ *Samsung Elecs. Am. Inc. v. Chung*, 325 F.R.D. 578,590 (N.D. Tex. 2017) (“A party resisting discovery must show how the requested discovery is overly broad, unduly burdensome, or oppressive by submitting affidavits or offering evidence revealing the nature of the burden”).

²⁶ FED. R. CIV. P. 26(b)(1).

²⁷ FED. R. EVID. 401 (“Evidence is relevant if: (a) it has any tendency to make a fact more or less probable than it would be without the evidence; and (b) the fact is of consequence in determining the action.”).

²⁸ *King v. Biter*, No. 15-cv-00414-LJO, 2021 U.S. Dist. LEXIS 116198, at *6 (E.D. Cal. July 25, 2017); *Marker v. Union Fid. Life Ins. Co.*, 125 F.R.D. 121, 125 (M.D.N.C. 1989) (“Conclusory claims of bad faith may not be the bases for conducting marginally relevant discovery which is by its nature burdensome. Such discovery requests amount to nothing more than an out of season fishing expedition.”).

²⁹ FED. R. CIV. P. 26(b)(1) advisory committee’s note to the 2000 amendment.

³⁰ *Id.*

³¹ *See Samsung Elecs. Am. Inc. v. Chung*, 321 F.R.D. 250, 280 (N.D. Tex. 2017) (“[T]he scope of discovery is not limited simply to ‘facts,’ but may entail other ‘matters’ that remain relevant to a party’s claims or defenses, even if not strictly fact-based.” (quoting *In re Blue Cross Blue Shield Antitrust Litig.*, No. 2:13-CV-20000-RDP, 2015 WL 9694792, at *2 (N.D. Ala. Dec. 9, 2015)).

Given this, it would not be surprising for courts to allow discovery of software that the legal department of a party might use.

If a piece of information is deemed relevant, it will still only be discoverable if the size of the request is “proportional to the needs of the case” under Rule 26(b)(1).³² After the 2015 amendments to the Rules, there was a renewed realization that all relevant information is not equally important.³³ Rule 26 lists proportionality “considerations,” which include the importance of the issues at stake, the amount in controversy, the parties’ relative access to relevant information, the parties’ resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit.³⁴ Courts have focused nearly exclusively on economic efficiency in their proportionality inquiries.³⁵ However, as TAR continues to proliferate and improve, the proportionality constraints built into the Federal Rules will be drained of much of their importance.³⁶ This is because TAR lessens the burden of production, the principle concern of the proportionality requirement, by automating what has traditionally been done manually.³⁷

III. PROTECTIONS FROM DISCOVERY

As the courts leaned into discovery to minimize information asymmetry, they realized the need for certain protections.³⁸ The first protection developed was the attorney-client privilege.³⁹ The purpose of this privilege is to encourage full and truthful communication between counsel and client,⁴⁰ and it is a privilege firmly grounded in public policy.⁴¹ However,

³² FED. R. CIV. P. 26(b)(1).

³³ Shaffer, *supra* note 13, at 234-35.

³⁴ *Id.* at 229.

³⁵ Endo, *supra* note 9, at 827.

³⁶ Engstrom & Gelbach, *supra* note 3, at 1051.

³⁷ See *infra* notes 74-80 and accompanying text for the discussion on the cost-reducing effects of TAR.

³⁸ Michael A. Blaise, *The Uncertain Foundation of Work Product*, 67 DePaul L. Rev. 35, 37 (2017).

³⁹ See Geoffrey C. Hazard Jr., *An Historical Perspective on the Attorney-Client Privilege*, 66 CALIF. L. REV. 1068 (1978).

⁴⁰ *E.g.*, *Upjohn Co. v. United States*, 449 U.S. 383, 389 (1981) (“Its purpose is to encourage full and frank communication between attorneys and their clients and thereby promote broader public interests in the observance of law and administration of justice. The privilege recognizes that sound legal advice or advocacy serves public ends and that such advice or advocacy depends upon the lawyer’s being fully informed by the client.”); see also *Fisher v. United States*, 425 U.S. 391, 403-04 (1976) (“Confidential disclosures by a client to an attorney made in order to obtain legal assistance are privileged. The purpose of the privilege is to encourage clients to make full disclosure to their attorneys.”).

⁴¹ See *Upjohn Co.*, 449 U.S. at 389; see also *Hunt v. Blackburn*, 128 U.S. 464, 470 (1888) (noting the privilege is based on the “necessity, in the interest . . . of justice, of the aid of persons having

only communications for the purpose of seeking or providing legal advice are protected by the privilege.⁴² This paper will not spend time on this privilege because its application is relatively straightforward.⁴³ Because attorney-client privilege covers communications between counsel and client, production requests asking for the input and output details surrounding legal technology will not typically implicate this protection. But if there are communications between attorney and client involved in the requested production, absolute protection applies to those communications.⁴⁴

A more interesting and controversial question looms regarding the application of the work-product protection. If legal tech is unevenly distributed and provides one side with a leg-up, then the other side may seek information surrounding those tools. Judges will need to determine whether the civil procedure rules should treat machine inputs and outputs in the same way as it treats traditional forms of attorney work-product.⁴⁵

The work-product doctrine generally shields attorneys from disclosing their case preparation materials, such as research memoranda and witness interviews.⁴⁶ It is a pragmatic doctrine that grew out of the recognition that attorneys serve both the advancement of justice *and* the interests of their clients.⁴⁷ It preserves the privacy of preparation that is crucial to attorney representation in an adversarial setting. With respect to the source of its authority, the work-product protection is a blend of codified procedure and judicial common law. Before *Hickman v. Taylor*, the 1947 Supreme Court case known to have established the doctrine, there was no rule regarding discovery protections.⁴⁸ As originally articulated by the Third Circuit, work-product was not its own doctrine, but rather a class of material protected by the discovery version of the attorney-client privilege.⁴⁹ It was a rule of public policy and was thought to encourage clients to make full disclosures to their

knowledge of the law and skilled in its practice, which assistance can only be safely and readily availed of when free from the consequences or the apprehension of disclosure”).

⁴² See *Fisher*, 425 U.S. at 403 (holding that the privilege protects “only those disclosures necessary to obtain informed legal advice which might not have been made absent the privilege”); see also *Diversified Indus., Inc. v. Meredith*, 572 F.2d 596, 602 (8th Cir. 1977) (“[T]he attorney must have been engaged or consulted by the client for the purpose of obtaining legal services or advice services or advice that a lawyer may perform or give in his capacity as a lawyer, not in some other capacity. A communication is not privileged simply because it is made by or to a person who happens to be a lawyer.”).

⁴³ See MATTHEW BENDER, 12 BENDER’S FORMS OF DISCOVERY TREATISE § 5.02 (2022).

⁴⁴ See, e.g., *Walker v. Cnty. of Contra Costa*, 227 F.R.D. 529, 536 (“Work product protection, unlike attorney client privilege, is not absolute. . .”).

⁴⁵ See Engstrom & Gelbach, *supra* note 3, at 1080.

⁴⁶ See *Hickman v. Taylor*, 329 U.S. 495, 510-11 (1947).

⁴⁷ *Id.*

⁴⁸ See Blaise, *supra* note 38, at 35-36.

⁴⁹ *Id.* at 46.

attorneys and encourage attorneys to “put their whole-souled efforts” into the case.⁵⁰

Following the Third Circuit’s articulation of the protection, the Supreme Court took up the issue in *Hickman v. Taylor*.⁵¹ After reinforcing the liberal discovery policy of the Federal Rules, the Court reflected on the effect that widespread intrusion into attorney work-product during discovery may have on the legal profession.⁵² It stressed that the proper preparation of a case requires assembling information, determining which facts are relevant, preparing legal theories, and planning strategy “without undue and needless interference.”⁵³ Thus, while there is always a burden that the requesting party show necessity in a discovery request, that burden is higher with attorney work-product.⁵⁴

Following the *Hickman* decision, significant confusion afflicted lower courts as to the application of the protection. In response, the Federal Rules Advisory Committee attempted to clarify things through an amendment in 1970.⁵⁵ It added a new rule that provided protection for “documents and tangible things that are prepared in anticipation of litigation” unless the requesting party can show that it has a “substantial need for the materials” in preparing its own case, and it cannot otherwise “obtain their substantial equivalent” without “undue hardship.”⁵⁶ With this language, the committee established the “substantial need” and “undue hardship” standards for a factual work-product⁵⁷ request and extended the protection to work by non-attorneys. This extension of the protection is important when thinking about technologies that are created and managed by non-attorney specialists.

The rule then goes on to say that, “if the court orders discovery of those materials, it must protect against disclosure of the mental impressions, conclusions, opinions, or legal theories” of the attorneys.⁵⁸ It affords added protection to an attorney’s “mental impressions, conclusions, opinions, or legal theories,” sometimes known as “opinion work-product.”⁵⁹ Most courts

⁵⁰ *Hickman v. Taylor*, 153 F.2d 212, 223 (3d Cir. 1945).

⁵¹ 329 U.S. 495 (1947).

⁵² Blaise, *supra* note 38, at 48-49.

⁵³ *Hickman*, 329 U.S. at 511.

⁵⁴ *See id.*

⁵⁵ Blaise, *supra* note 38, at 52-53.

⁵⁶ FED. R. CIV. P. 26(b)(3)(A).

⁵⁷ *See supra* notes 53-55 and accompanying text for distinction of factual work-product versus opinion work-product.

⁵⁸ FED. R. CIV. P. 26(b)(3)(B) (“If the court orders discovery of those materials, it must protect against disclosure of mental impressions, conclusions, opinions, or legal theories of a party’s attorney or other representative concerning the litigation.”).

⁵⁹ Blaise, *supra* note 38, at 54.

have held that opinion work-product is never discoverable, elevating it to a near absolute privilege.⁶⁰ Work-product that is not included in opinion work-product is referred to as “fact work product.”⁶¹ Fact work-product is governed by the “substantial need” and “undue hardship” standards in subsection (A) of the rule and can be discoverable.⁶² Given the text of the rule, fact work-product that falls in subsection (A) only includes “documents and tangible things.”⁶³ This means that, textually, intangible work-product is only protected if it falls into subsection (B), the one that provides absolute protection for opinion work-product.⁶⁴

The work-product doctrine is one of the fuzzier tools in the discovery toolkit and its scope is hotly debated. Despite this, it is still widely recognized as a pillar of a deeply adversarial model of law rooted in competition.⁶⁵ The doctrine creates a “zone of privacy” within which counsel can operate free of interference and without worry that outputs will fall into the others’ hands, permitting counsel to focus on zealous client representation. It also protects against free riding on an adversary’s diligence.⁶⁶ Work-product protection creates the conditions necessary for a well-functioning adversarial system by safeguarding returns on, and thus investment in, legal talent.⁶⁷ In the same way, it may serve to protect investment in legal technology. These purposes are important to keep in mind as we examine its applicability to emerging developments.

IV. APPLYING PRINCIPLES OF DISCOVERY TO EMERGING TECHNOLOGY

As judges apply their discretion to address discovery disputes surrounding emerging technologies, they should keep in mind the overarching principles and values embodied in common law and the Federal

⁶⁰ See *In re* Murphy, 560 F.2d 326, 336 (8th Cir. 1977) (“[O]pinion work product enjoys a nearly absolute immunity and can be discovered only in very rare and extraordinary circumstances.”); *Holmgren v. State Farm Mut. Auto. Ins. Co.*, 976 F.2d 573, 577 (9th Cir. 1992) (requiring “a compelling need” to obtain material otherwise considered opinion work-product); *Duplan Corp. v. Moulinage et Retorderie de Chavanoz*, 509 F.2d 730, 734 (4th Cir. 1974) (“[N]o showing of relevance, substantial need or undue hardship should justify compelled disclosure of an attorney’s mental impressions, conclusions, opinions or legal theories.”).

⁶¹ *In re* Grand Jury Proc. #5, 401 F.3d 247, 250 (4th Cir. 2005).

⁶² FED. R. CIV. P. 26(b)(3)(A) (“Ordinarily, a party may not discover documents and tangible things that are prepared in anticipation of litigation. . . . But, subject to Rule 26(b)(4), those materials may be discovered if: (i) they are otherwise discoverable under Rule 26(b)(1); and (ii) the party shows that it has substantial need for the materials to prepare its case and cannot, without undue hardship, obtain their substantial equivalent by other means.”).

⁶³ *Id.*

⁶⁴ FED. R. CIV. P. 26(b)(3)(B) (see *supra* note 58 for full text of subsection).

⁶⁵ Engstrom & Gelbach, *supra* note 3, at 1079.

⁶⁶ *Id.* at 1077.

⁶⁷ *Id.* at 1077-78.

Rules. The challenge for the courts, as well as lawmakers, is in applying the concerns embedded in the discovery process and protections, which originated in a very different era.⁶⁸ Practitioners, too, ought to be interested in knowing what to expect regarding their tech tools' exposure to discovery requests. Practitioners may also want to know what they may be able to obtain from their adversaries.

Ultimately, in deciding whether to compel discovery upon motion from one of the parties, district judges should exercise their discretionary authority in light of several considerations that can be gleaned from the federal rules laid out in the prior two parts. Courts ought to ask themselves the following questions:

First, is the subject of the tech-related request relevant to proving or disproving the underlying claims of the litigation? If it is not, is it relevant to the adequacy of a discovery production? Only requests that are either relevant to the merits of the dispute, so called merit-directed discovery, or are reasonably necessary for assessing the quality of a production, so called process-directed discovery, should be compelled.⁶⁹

Second, what type of "attorney work" is involved? Does the requested production pose a significant risk of exposing attorney mental impressions, or does it primarily involve the rote input of external factors and data? Where there is attorney labor or judgment, absolute protections should be afforded. A lighter and more qualified protection is owed to mere facts collected by the attorney.⁷⁰

Third, during which point(s) in the lifecycle of the development and usage of the relevant legal technology was attorney skill and labor utilized? Put another way, was the attorney's skill utilized purely in the initial development of the technology (unrelated to a lawsuit) or in the moment the machine was used for the present litigation? Or was it only after the technology had served its role that the attorney added his or her opinion? For attorney work to be protected by the work-product doctrine, it must have been created in anticipation of or during litigation.

A. *Technology Assisted Review ("TAR")*

Today's civil discovery is increasingly electronic discovery ("e-discovery") because, at least for the past couple of decades, the vast majority of all information has been stored in electronic form (ESI), and that share

⁶⁸ Engstrom & Gelbach, *supra* note 3, at 1086.

⁶⁹ See *supra* notes 13-15 and accompanying text for the discussion on "merit-directed discovery" versus "process-directed discovery."

⁷⁰ See *supra* notes 57-62 and accompanying text for the discussion on "opinion work-product" versus "fact work-product."

gets higher each year.⁷¹ This proliferation of ESI is due to the decreases in data storage costs and increases in processor speed.⁷² It has led to an increase in the overall quantity of data stored and has resulted in a more expansive and burdensome review process.⁷³ This is where TAR, which seeks to reduce the extent of human involvement in the e-discovery process, comes in.⁷⁴ It is the process by which computers assist in searching a database for keywords that have been deemed relevant.⁷⁵ TAR is typically conducted in two stages: a first-pass review where a machine identifies likely responsive or relevant documents, and a second-pass usually involving a human that checks for the application of any privileges or protections.⁷⁶

Some are hesitant to embrace TAR because it is imperfect to a degree;⁷⁷ however, numerous studies establish that well-implemented TAR tools are as good as, and often better than, purely human review.⁷⁸ So despite some lingering uncertainty about its performance, TAR is accepted by much of the judiciary⁷⁹ and has been viewed by some commentators as one of the best

⁷¹ Daniel Martin Katz, *Quantitative Legal Prediction—or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L.J. 909, 943 (2012).

⁷² *Id.*

⁷³ *Id.* at 943-44 (explaining how conducting an exhaustive manual review of ESI has become nearly impossible, thereby requiring lawyers to seek alternative approaches to reviewing and producing information).

⁷⁴ *Id.* at 945.

⁷⁵ McGinnis & Pearce, *supra* note 4, at 3047. Parties may come to an agreement during the Rule 26(f) conference at the beginning of discovery regarding what is relevant, or a party may make this deliberation on its own. Engstrom & Gelbach, *supra* note 3, at 1045 (explaining that with input from lawyers and computer technicians, TAR uses machine learning classifiers to flag relevant and privileged documents).

⁷⁶ Ralph C. Losey, *Predictive Coding and the Proportionality Doctrine: A Marriage Made in Big Data*, 26 REGENT U. L. REV. 7, 55-56 (2013) (explains that the first pass through is programmed by humans and given a seed set that the machine uses to identify relevancy. The second pass does not consider any documents rejected in the first-pass review, but it does scrutinize the output of the machine for privileged material before anything is disclosed.).

⁷⁷ McGinnis & Pearce, *supra* note 4, at 3047.

⁷⁸ Engstrom & Gelbach, *supra* note 3, at 1052 (“[This is] in terms of recall (i.e., the proportion of documents in the total pool of documents that the tool accurately identifies as relevant) and . . . precision (i.e., the proportion of documents among those that the tool identifies that are in fact relevant”).

⁷⁹ *See, e.g.*, *Moore v. Publicis Groupe*, 287 F.R.D. 182, 187, 193 (S.D.N.Y. 2012) (“[TAR] works better than most of the alternatives, if not all of the [present] alternatives” and “[c]omputer-assisted review now can be considered judicially-approved for use in appropriate cases.”); *Nat’l Day Laborer Org. Network v. U.S. Immigr. & Customs Enf’t Agency*, 877 F. Supp. 2d 87, 109 (S.D.N.Y. 2012) (“[P]arties can (and frequently should) rely on . . . machine learning tools to find responsive documents.”); *Dynamo Holdings Ltd. P’ship v. Comm’r*, 143 T.C. 183, 192 (2014) (“[W]e understand that the technology industry now considers predictive coding to be widely accepted for limiting e-discovery to relevant documents and effecting discovery of ESI without an undue burden.”); *Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125, 126 (S.D.N.Y. 2015) (“[TAR] is an acceptable way to search for relevant ESI in appropriate cases.” (quoting *Moore*, 287 F.R.D. 182, 183 (S.D.N.Y. 2012))). *But see* *Progressive Cas. Ins. Co. v.*

tools for achieving proportionality because of its cost-reducing effects.⁸⁰ Courts do not yet mandate its use, though.⁸¹

In any case, as TAR's use continues to proliferate, courts will be forced to address whether a producing party should be compelled to disclose its TAR methodology. Adversaries have already begun requesting "discovery on discovery," or discovery of the process by which a party produced the requested information.⁸² There are two types of disclosures that might be requested regarding a production that was made using TAR: the seed sets and other information used to train the machine, or the search terms and other inputs used to make a specific production. Courts try to rely on party cooperation with these requests where possible, but when conflicts arise, they must intervene.⁸³ Lower courts are grappling with how much inter-party cooperation to require when implementing TAR protocols.⁸⁴ Some strongly encourage disclosure but do not mandate it.⁸⁵ Others have required disclosure or make it a condition of a party's use of TAR.⁸⁶ For example, the Northern District of Illinois in *In re Boiler Chicken Antitrust Litigation* issued an ESI order that required the parties to disclose, among other things, how the TAR process would work, including whether software accounted for common misspellings and/or synonyms.⁸⁷

Delaney, No. 2:11-cv-00678-LRH-PAL, 2014 WL 3363467, at *11 (D. Nev. July 18, 2014) (refusing party request to use TAR but mostly because of the party's bad faith in abandoning an agreed-to protocol).

⁸⁰ Losey, *supra* note 76, at 54.

⁸¹ See, e.g., *Hyles v. New York City*, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016) ("[I]t is not up to the Court, or the requesting party . . . to force [the defendant] . . . to use TAR when it prefers to use keyword searching."); *In re Viagra (Sildenafil Citrate) Prods. Liab. Litig.*, 2016 U.S. Dist. LEXIS 144925, at *52-53 (N.D. Cal. Oct. 14, 2016) ("Even if predictive coding were a more efficient and better method, . . . it is not clear on what basis the [c]ourt could compel [defendant] to use a particular form of ESI, especially in the absence of any evidence that [defendant's] preferred method [of using search terms] would produce, or has produced, insufficient discovery responses.").

⁸² Engstrom & Gelbach, *supra* note 3, at 1080. This is also referred to as "process-directed discovery."

⁸³ See, e.g., *Mixing Equip. Co. v. Innova-Tech, Inc.*, 1987 U.S. Dist. LEXIS 10107, at *3-4 (E.D. Pa. Mar. 14, 1987).

⁸⁴ Engstrom & Gelbach, *supra* note 3, at 1058.

⁸⁵ See, e.g., *Aurora Coop. Elevator Co. v. Aventine Renewable Energy-Aurora W., LLC*, 2015 WL 10550240, at *2 (D. Neb. Jan. 6, 2015) (clarifying that Rule 26(b)(1) does not authorize discovery of documents in a seed set but then encouraging cooperation between the parties); *In re Biomet M2a Magnum Hip Implant Prods. Liab. Litig.*, 2013 WL 6405156, at *2 (N.D. Ind. Aug. 21, 2013) (mem.) (noting lack of authority under Rule 26(b)(1) but urging the party to re-think its refusal to produce); *Winfield v. City of New York*, 2017 WL 5664852, at *12 (S.D.N.Y. Nov. 27, 2017).

⁸⁶ *Progressive Cas. Ins. Co. v. Delaney*, 2014 WL 3563467, at *11 (D. Nev. 2014) (refusing a request to use TAR because the party advocating it had refused to share seed sets and other methodological details); *Fed. Hous. Fin.-Agency v. HSBC N. Am. Holdings Inc.*, 2014 WL 584300 (S.D.N.Y. 2014); *Moore v. Publicis Groupe*, 287 F.R.D. 182, 199 (S.D.N.Y. 2012) (ordering collaboration, including "iterative seed selection" and "quality control processes").

⁸⁷ 2018 U.S. Dist. LEXIS 33140, at *34-35 (N.D. Ill. Jan. 3, 2018).

Many federal courts find “discovery on discovery” to sometimes be permissible because information about a discovery production can aid a party in the preparation of their case and fall within the relevance standard.⁸⁸ Courts are most open to this type of discovery when there are serious questions about the sufficiency of a merit-based discovery production.⁸⁹ Discovery concerning the search and maintenance of an opposing party’s information systems might be “relevant” where the requesting party can “point to the existence of additional responsive material” or when the documents already produced “permit a reasonable deduction that other documents may exist or did exist and have been destroyed.”⁹⁰ A party moving to compel process-directed discovery should be prepared to substantiate their claims of relevance and proportionality with factual support.⁹¹ For these reasons, many courts often reject work-product defenses to the disclosure of a seed set used to train a TAR model.⁹² Seed sets refer to the input data that are used to “train” a machine to recognize specific types of data based on the instructions it has been given.⁹³ Additionally, a set of decisions from West Virginia courts found that search terms used for TAR did not constitute work-product.⁹⁴ Search terms refer to the words and phrases chosen by counsel that will be used to comb through a large set of documents and data, isolating those with the terms.⁹⁵

⁸⁸ See, e.g., *Ruiz-Bueno v. Scott*, 2013 U.S. Dist. 162953, at *6 (N.E. Ohio Nov. 5, 2013) (reasoning that when “information about discovery is a matter which may “aid a party in the preparation . . . of his case,” it falls within the relevance standard of Rule 26(b)(1)).

⁸⁹ See, e.g., *Cannata v. Wyndham Worldwide Corp.*, 798 F. Supp. 2d 1165 (D. Nev. 2011) (an inability to obtain process-directed information might make it difficult, if not impossible, to evaluate the reasonableness and thoroughness of a party’s efforts to search for and produce relevant ESI); *Burnett v. Ford Motor Co.*, 2015 U.S. Dist. LEXIS 48623 (S.D. W. Va. Apr. 14, 2015) (in the face of questions on the adequacy of a production, common sense dictates that the party conducting the search must share information regarding the universe of potentially relevant documents being preserved, and those that no longer exist, as well as the search terms used in collecting relevant documents. The court held that the party responsible for the search and production has the duty to demonstrate its reasonableness).

⁹⁰ *Hubbard v. Potter*, 247 F.R.D. 27 (D.D.C. 2008).

⁹¹ *Ford Motor Co. v. Edgewood Props.*, 257 F.R.D. 418, 427 (D.N.J. 2009) (although the court did not preclude the possibility of process-directed discovery, it left unresolved what would constitute a “colorable showing” sufficient to support the defendant’s request); *Orillaneda v. French Culinary Inst.*, 2011 U.S. Dist. LEXIS 105793 (S.D.N.Y. Sept. 19, 2011) (held that process-directed discovery is not available absent some factual showing of a production deficiency).

⁹² See *Hinterberger v. Cath. Health Sys., Inc.*, 2013 WL 2250591, at *22 (W.D.N.Y. May 21, 2013) (reasoning that, neither the scanning nor objective coding work required any access or need for confidential information, including work-product).

⁹³ *Engstrom & Gelbach*, *supra* note 3, at 1015.

⁹⁴ See *Burd v. Ford Motor Co.*, 2015 WL 4137915, at *10 (S.D.W.Va. July 8, 2015) (holding that the sharing of search terms used by custodians who searched the defendant’s records does not amount to attorney work-product); *Burnett v. Ford Motor Co.*, 2015 WL 4137847, at *10 (S.D.W.Va. July 8, 2015) (finding that search terms and custodian names are not work-product).

⁹⁵ *Engstrom*, *supra* note 3, at 1015.

Courts making these determinations also consider the risk of conveying counsel’s mental impressions or revealing other strategically valuable information. With search terms, many conclude there is not a substantial risk because of the sheer amount of material that is typically produced using the selected terms.⁹⁶ However, the process of selecting a seed set arguably reveals attorney reasoning and understandings of the case and so justifies the fuller protection of opinion work-product.⁹⁷

Another camp of federal courts are those that are wary of allowing process-directed discovery, especially early on in a case.⁹⁸ Some have held that discovery deficiencies, standing alone, are insufficient to warrant process-directed discovery; they say the proper analysis is a balancing test that weighs the burdens of the additional discovery against the likely benefits of it.⁹⁹ Many litigants argue process-directed discovery should not be allowed except under narrow circumstances because it can result in unnecessary expenses and, in any case, *how* a party chooses to comply with its discovery obligations should not be relevant.¹⁰⁰ For example, the Southern District of New York in *Winfield v. City of New York* refused to compel disclosure, holding that “in the absence of evidence of good cause,” there is no basis for “courts to insert themselves as super-managers of the parties’ internal review processes, including training of TAR software, or to permit discovery about such process[.]”¹⁰¹

Courts should make these determinations by looking at the circumstances of the technology and case in front of them through the lens of the Federal Rules and work-product doctrine. We return to the framework

⁹⁶ See *Disability Rts. Council of Greater Wash. v. Wash. Metro. Transit. Auth.*, 242 F.R.D. 139, 141-44 (D.D.C. 2007) (“With the number of those documents said to be totaling into the thousands, it would be difficult to convenience that Plaintiff’s trial strategy could be gleaned solely by . . . disclosure of the documents selected”); In re *Shell Oil Refinery*, 125 F.R.D. 132, 134 (E.D.La. 1989) (“[I]t is highly unlikely that defendant will be able to discern the plaintiff’s theory of the case or thought processes simply by knowing which 65,000 documents out of 660,000 documents have been selected for copying”).

⁹⁷ See, e.g., *Shelton v. Am. Motors Corp.*, 805 F.2d 1323, 1328-29 (8th Cir. 1986) (holding document selection is protected as work-product because counsel “identified, selected, and compiled documents that were significant to her client’s defenses in this case”); In re *Allen*, 106 F.3d 582, 608 (4th Cir. 1997) (concluding that counsel’s choice of materials constituted work-product); *Spork v. Peil*, 759 F.2d 312, 316 (3d Cir. 1985) (holding counsel’s selection of documents for deposition protected work-product because the “process of selection and distillation” can “reveal important aspects of [an attorney’s] understanding of the case”).

⁹⁸ See, e.g., *Miller v. York Risk Servs. Grp.*, 2013 U.S. Dist. LEXIS 173354 (D. Ariz. Dec. 9, 2013); *Martin v. Allstate Ins. Co.*, 416 N.E.2d 347 (1st Dist. 1981) (The plaintiff was not permitted to address during the defendant’s deposition topics such as “document retention policies,” “defendant’s efforts in responding to plaintiff’s discovery,” or “defendant’s efforts . . . to preserve documents and electronic information relevant to the anticipated suit”).

⁹⁹ *Larsen v. Caldwell Banker Real Est. Corp.*, 2012 U.S. Dist. LEXIS 12901 (C.D. Cal. Feb. 2, 2012).

¹⁰⁰ Shaffer, *supra* note 13, at 215.

¹⁰¹ *Winfield v. City of New York*, 2017 U.S. Dist. LEXIS 194413, at *10 (S.D.N.Y. Nov. 27, 2017).

of considerations laid out earlier in this paper: (1) the relevancy of the requested material, either to the merits or the adequacy of production; (2) the risk of exposing attorney mental impressions; and (3) whether the involved attorney exercised his or her judgment in anticipation of litigation.

With respect to relevancy, the seed set or search terms used for a TAR program will not be relevant to the underlying merits of the dispute because they exist solely in the context of the discovery proceedings. However, if the judge finds that a production regarding the implementation of a TAR program is genuinely needed to evaluate the reasonableness and thoroughness of a party's efforts to search for and produce ESI relevant to the merits of the dispute, then a limited discovery is warranted. This argument, that "discovery on discovery" is necessary to evaluate the adequacy of a production, applies to seed sets depending on the circumstances and is almost always applicable to a set of search terms, which is directly relevant to the thoroughness of a production. Ideally, these terms are agreed to by the parties ahead of time, but this may not always happen.

The second important question in the evaluation is whether counsel's mental impressions and understandings of the case are revealed through a disclosure of seed sets or search terms for a TAR program. With seed sets created for a particular litigation, there is significant work and deliberation involved in the development of the program. It requires humans to instruct it, to train its models and regularly refine them over time.¹⁰² During the refinement process, the lawyer chooses an "active learning protocol" to select further training documents to add to the seed set. It's a process that requires deliberation and expertise. This pre-prediction process is arguably one form of a mental impression. On the other hand, search terms are typically relevant factual terms that correspond to the production request made by the other party; it is questionable as to whether attorney mental impressions are exposed through their selection. Even if there were mental impressions to some degree involved in the search term selection, they will often be outweighed by a need to assess the sufficiency of a production.

Even if a seed set is deemed to contain mental impressions of an attorney, it should only be protected by the work-product protection if it was created in anticipation of litigation, which is the third consideration in

¹⁰² Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 28 (2017)(Machine learning models are not just turned loose on data, rather, programmers make numerous decisions about how to partition data, which model types and data features to choose, and how much to tune the model); Katz, *supra* note 71, at 946-53 (The task is to use a trained model to classify new records relative to the gold standard data, the seed set, that has been preidentified or pre-classified by an expert reviewer); Engstrom, *supra* note 3, at 1046 (The lawyers perform traditional document review on a subset of a production to create a seed set to train the model, then they engage in further such efforts as the system iteratively moves toward a best model).

evaluating the application of the protection. While, at some point in the process, the team of attorneys and non-attorneys developing TAR tools do work and leave their mental impressions, it is not so clear that these mental impressions are expressed in anticipation of litigation. Legal teams may develop a seed set for a TAR program well in advance of any particular litigation, or one may be developed for the particular case at hand. The degree to which disclosure is necessary to assess the sufficiency of a production may also vary. Given the discretion of the district courts, this decision will come down to a case-by-case assessment. While search terms will almost always satisfy this third consideration because they are developed for a specific dispute, they will typically not be protected because of their high relevance to the adequacy of a production, and thus they do not clear the hurdle of the first consideration.

Given the uncertainty of how these “discovery on discovery” disputes will be handled by the courts, litigants are best off conferring early on and frequently with opposing counsel throughout the pre-trial process so they can resolve things in as mutually beneficial a way as possible and avoid unnecessary litigation costs.

B. Predictive Analytics

Outcome forecasting using data analytic tools is one of the less developed forms of AI in the legal sector. However, it is growing quickly¹⁰³ and will become more useful and accessible as data collection becomes cheaper and technology advances. Data analytics can be used to forecast or predict outcomes at various stages of litigation and in multiple settings. Predictions can be jurisdiction- or judge-specific and used to compare forums and assess case quality at intake, filing, or once litigation is underway. They can inform litigation and settlement strategy by giving likely damages, the likely ruling of the assigned judge, and the likely timeframe for resolution. These tools can also be used to analyze judge-level data to tailor arguments to the bench.¹⁰⁴ One example of predictive analytics in the legal market is Lex Machina, which uses data mining and predictive analytics techniques to forecast litigation outcomes.¹⁰⁵ Another example is LexPredict, a company that has built models to predict the outcome of U.S. Supreme Court cases at accuracy levels challenging experienced Supreme

¹⁰³ Engstrom, *supra* note 3, at 1059 (“Surveys suggest substantial recent increases in use of data-based outcome prediction tools among law firms”); G. Patrick Flanagan & Michelle Hook Dewey, *Where do We Go from Here: Transformation and Acceleration of Legal Analytics in Practice*, 35 GA. ST. U. L. REV. 1245, 1253 (2019) (“Predicting court behavior is arguably the fastest growing sector of the legal-analytic marketplace”).

¹⁰⁴ Engstrom, *supra* note 3, at 1011.

¹⁰⁵ Michael Mills, *Using AI in Law Practice: It's Practical Now*, 42 L. PRAC. 48, 50 (2016).

Court practitioners.¹⁰⁶ The company Premonition says it is using data mining and other AI techniques “to expose, for the first time ever, which lawyers win the most before which judges.”¹⁰⁷

Predictive analytics works by using machine learning to process historical data, identify patterns, and assess the likelihood of specific outcomes.¹⁰⁸ In other words, “known data can be used to predict what will happen in situations that have not yet occurred.”¹⁰⁹ Whether generated by experience or a sophisticated algorithm, prediction is a core component of the guidance that most lawyers offer. A lawyer’s job at its core is generating informed answers to client questions. However, “when it comes to processing and deriving insights from large-scale data or document sets, humans have important cognitive limitations.”¹¹⁰ “Even if one has access to all of the relevant information, without the aid of technology . . . , it is essentially impossible to completely process all relevant data or its potentially relevant dimensions.”¹¹¹ Predictive technologies are designed to remedy or supplement human shortcomings through the quicker, cheaper, and unbiased processing of large sets of data.¹¹²

A lawyer’s analysis and evaluation inform a client’s resolution of his or her dispute, and it is fair to say that better litigation decisions can be made with greater information about judges, opposing counsel, causes of action, and procedural posture.¹¹³ Data points surrounding these considerations can be gathered from dockets and other collections of historical data. Anecdotal and personal experiences can be quantified, transferred, and utilized by attorneys foreign to the jurisdiction.¹¹⁴ Outcome prediction is involved in multiple parts of the litigation process. For example, the decision whether to originate a litigation matter and, if so, where, requires a delicate balancing

¹⁰⁶ *Id.*

¹⁰⁷ *Id.*

¹⁰⁸ Rhys Dipshan, *With Analytics Tools, Law Firms Are Adding Predictive Power to Their Advice*, THE AM. LAW. ONLINE (June 3, 2022), <https://www.law.com/americanlawyer/2022/06/03/with-analytics-tools-law-firms-are-adding-predictive-power-to-their-advice/> [<https://perma.cc/CB7E-K29Y>]; For a deeper explanation of predictive analytic technologies, see Katz, *supra* note 71.

¹⁰⁹ McGinnis & Pearce, *supra* note 4, at 3052.

¹¹⁰ Katz, *supra* note 71, at 929.

¹¹¹ *Id.*

¹¹² *Id.* at 928. A human reasoner’s “understanding of likelihood might be driven by personal observations that are anecdotal, censored, or otherwise not indicative of the true distribution of outcomes.”

¹¹³ Mark K. Osbeck, *Lawyer as Soothsayer: Exploring the Important Role of Outcome Prediction in the Practice of Law*, 123 PENN. ST. L. REV. 41, 43 (2018) (“One of the most important tasks lawyers undertake in furtherance of this advisory role is outcome prediction: that is, advising the client as to the likely outcome of various legal proceedings”).

¹¹⁴ Flanagan & Dewey, *supra* note 103, at 1254.

of costs, risks, benefits, and likelihoods of success.¹¹⁵ Additionally, deciding whether to accept a settlement offer and for how much involves an assessment of the likely outcome in the absence of a settlement.¹¹⁶ As these technologies are incorporated more into pre-trial decisions, parties will become more curious to know the factors going into their adversary's predictions.

One use of outcome prediction technology is in forum shopping. While plaintiffs have "venue privilege" (the ability to make the first court selection), defendants may move for statutory transfer to a new district or use a *forum non conveniens* motion.¹¹⁷ The result is that litigants on both sides have a say in where a case is adjudicated, and both will likely have an interest in a particular forum.¹¹⁸ After having used legal technology to assess the probabilities of success in multiple available forums, plaintiffs will be able to choose to file their suit in the one most favorable to their client, while defendants can move to transfer or dismiss.

Another use can be seen in the settlement context, where likelihoods of success at trial carry substantial weight in the decision whether to settle, and if so, at what amount. For example, in 2020, Schiff Hardin (prior to its merger with Arent Fox) developed an internal analytics tool to predict outcomes for a client's claim prior to settlement.¹¹⁹ Baker McKenzie, too, has launched a partnership with SparkBeyond to predict the likelihood of future risks for clients.¹²⁰ With this information, the tech user has a significant advantage in negotiations. There are also several legal tech providers, such as Lex Machina, Wolters Kluwer and Trellis, that offer predictive analytics tools to in-house clients. Their tools offer predictions around litigation outcomes, outside counsel costs, case timelines, and more.¹²¹

These uses of predictive analytics raise questions about whether judges should be "empowered to order parties to disclose their machine outputs" or even ensure that the parties have access to the same prediction tools.¹²² Least aggressively, courts could require the parties to disclose the fact of their use

¹¹⁵ Osbeck, *supra* note 112, at 46.

¹¹⁶ *Id.*

¹¹⁷ *See, e.g.*, *Atl. Marine Constr. Co. v. United States Dist. Court*, 571 U.S. 49 (2013).

¹¹⁸ Engstrom & Gelbach, *supra* note 3, at 1060-61. There are cost and convenience considerations based on location and time length of the suit, for example. The choice of forum can also affect which law applies and influence the jury pool. Judges in some jurisdictions might be more plaintiff friendly than others or view certain motions more favorably.

¹¹⁹ Dipshan, *supra* note 108.

¹²⁰ *Id.* ("When Baker McKenzie launched its innovation program, Reinvent, in October 2020 in collaboration with software provider SparkBeyond, for instance, it also announced it is leveraging the partnership to predict the likelihood of future risks for clients").

¹²¹ *Id.*

¹²² Engstrom & Gelbach, *supra* note 3, at 1007.

of predictive analytics. More aggressively, parties could be required to disclose to all sides, including the judge, their machines' predictions. Most aggressive of all would be a requirement that a party who uses predictive analytics give direct access to the programs or code used to generate predictions.¹²³

Courts must think about whether the decision guidance provided by these tools should be protected. They should make these determinations by looking at (1) the relevancy of the requested material either to the merits of the case or the adequacy of production; (2) the risk of exposing attorney mental impressions if the disclosure is compelled; and (3) whether the attorney judgment involved was exercised in anticipation of litigation.¹²⁴

It is important to keep in mind the core focus of the work-product doctrine: attorney labor and mental impressions in anticipation of litigation.¹²⁵ It can reasonably be argued that a tool that requires no more than the attorney to feed in the pleadings and papers to date, or a tool primed by inputting only a set of basic case facts, does not involve substantial lawyerly mental impressions or work. In other words, a judge is unlikely to find a machine output that compares the likelihood of success in one forum versus another based on external factors (such as docket loads, or the characteristics of the judge and jury) as the kind of information production that the work-product rule is designed to protect.¹²⁶

However, even if it is not in the moment of prediction, there are attorney work and mental impressions that go into developing predictive analytic tools that accurately forecast legal outcomes. These intelligent machines require humans to train their models and regularly refine them over time.¹²⁷ Both the initial training and the maintenance of an up-to-date predictive tool requires people that understand both computer programming and the underlying legal tasks. Most of these tools use machine learning methods that analyze a set of "training data," or "human-labeled data inputs," in order to "draw predictive inferences about the labels humans would assign to new and unseen instances."¹²⁸ "One needs to be able to 'pre-predict' the set of cases that are sufficiently similar to the base case to be indexed for purposes

¹²³ *Id.* at 1070 ("Disclosure could, in turn, lead to crafting of new rules, whether by judges or via the rulemaking process, distinguishing types of reasons surfaced via predictive analytics.").

¹²⁴ See *supra* notes 68-70 and accompanying text.

¹²⁵ See *supra* notes 55-64 and accompanying text.

¹²⁶ *Id.* at 1085.

¹²⁷ See Desai & Kroll, *supra* note 102 (Machine learning models are not just turned loose on data, rather, programmers make numerous decisions about how to partition data, which model types and data features to choose, and how much to tune the model); Katz, *supra* note 71, at 953 (the goal is "to optimally update the model automatically as time ticks forward.").

¹²⁸ Engstrom & Gelbach, *supra* note 3, at 1015.

of executing the actual prediction of case outcomes.”¹²⁹ This task involves significant discretion on the part of the lawyer and requires experience and expertise that may be worth protecting during discovery.

Investments in legal technology may also be worth protecting during discovery. For starters, access to good data is a prerequisite to building an accurate prediction system. Good data needs to be consistent, detailed, and sufficiently large to be statistically valid.¹³⁰ The legal field makes it difficult to find reliable, centralized data sets because information must be sourced from different entities and levels, such as federal, state, county, and city laws.¹³¹ The complexity of this process makes law firms much better off if they partner with a data science company, at least for the development of the program.¹³² All of this makes good data tough to collect and therefore expensive. These investments should be considered for protection from discovery.

The issue with this form of work and investment is that the work-product protection only applies to work done in anticipation of litigation.¹³³ It could be argued that the training of the models well before any litigation is still intended to be used in litigation, especially given the Supreme Court has held that work-product protection can extend beyond the specific litigation for which the materials were prepared.¹³⁴ Lower courts have interpreted authoritative case law to mean that “a document was prepared ‘in anticipation of litigation’ if it was prepared or obtained ‘because of’ the prospect of litigation.”¹³⁵ However, they vary in their application of this principle.¹³⁶ Many of these programs are designed to counsel clients in litigation generally, but may not have been created in the midst of or in

¹²⁹ Katz, *supra* note 71, at 955-57 (“In order to deliver optimal results, the retrieval or “pre-predicted” set of comparison cases needs to include cases that share an analogical structure to the reference case. The key research-and-development challenge is to develop a refined, but also scalable, method for defining similarity.”).

¹³⁰ Mills, *supra* note 105, at 50.

¹³¹ *Id.*

¹³² Dipshan, *supra* note 108 (“It is difficult for law firms to successfully develop such tools alone,” says Danielle Benecke, co-founder of Baker McKenzie’s machine learning practice. She notes that the firm’s partnership with SparkBeyond provided the machine learning building blocks to help the firm develop internal data science expertise.”).

¹³³ FED. R. CIV. P. 26(b)(3)(A).

¹³⁴ See *F.T.C. v. Grolier Inc.*, 462 U.S. 19, 30-31 (1983).

¹³⁵ Engstrom, *supra* note 3, at n.315.

¹³⁶ See *United States v. Adlman*, 134 F.3d 1194, 1202 (2d Cir. 1998) (holding document must “have been prepared or obtained because of the prospect of litigation”); *In re Sealed Case*, 146 F.3d 881, 884 (D.C. Cir. 1998) (“[T]he lawyer must have had a subjective belief that litigation was a real possibility, and that belief must have been objectively reasonable.”); *United States v. Davis*, 636 F.2d 1028, 1040 (5th Cir. 1981) (applying work-production protection only where the “primary motivating purpose behind the creation of the document was to aid in possible future litigation”).

anticipation of a particular case.¹³⁷ As a side note, while much of the preparing of these tools is done by non-attorney specialists, that fact alone does not exclude them from work-product protection.¹³⁸

Furthermore, once the predictive program provides its forecasts, there is still significant work left for the lawyer. The machine outputs will almost always serve as just one factor in the lawyer's next decision because the outputs are imperfect and are likelihoods rather than certainties.¹³⁹ Strategic insights "must account for gaps in coverage" and should entail consulting several sources due to the "current actuality of market offerings and restrictions."¹⁴⁰ Any follow up strategizing, editing, or mental impressions that the lawyers express after the machine gives its outputs should always be protected as work-product.

At first glance, a predictive analytic tool might at least qualify only for Rule 26(b)(3)(A)'s lower qualified fact work-product protection.¹⁴¹ Arguing a motion to compel disclosure would require a showing of a "substantial need" for the information and "undue hardship" in obtaining its equivalent elsewhere.¹⁴² Courts heavily scrutinize claims of inconvenience and hardship, often requiring a demonstration that it is "significantly more difficult, time-consuming or expensive to obtain the information from another source."¹⁴³ However, a party can make a necessity showing where a deep pocketed, repeat player enjoys privileged access to data, making replication of an analysis nearly impossible if not extremely burdensome.¹⁴⁴

¹³⁷ Engstrom, *supra* note 3, at 1084-85; *See* Prater v. Consol. Rail Corp., 272 F. Supp. 2d 706, 713 (N.D. Ohio 2003) (holding that a study of employee repetitive stress complaints, although performed at counsel's direction, were business and not legal work, despite being motivated by past lawsuits and risk of future lawsuits); *United States v. Textron Inc. & Subsidiaries*, 577 F.3d 21, 29-31 (1st Cir. 2009) (holding that documents created in the ordinary course of compliance with auditor and securities filing requirements were not created because of litigation).

¹³⁸ *United States v. Nobles*, 422 U.S. 225, 238-39 (1975).

¹³⁹ McGinnis, *supra* note 4, at 3053.

¹⁴⁰ Flanagan, *supra* note 103, at 1258.

¹⁴¹ Engstrom, *supra* note 3, at 1082.

¹⁴² FED. R. CIV. P. 26(b)(3)(A).

¹⁴³ *Gucci Am., Inc. v. Guess?, Inc.*, 271 F.R.D. 58, 81 (S.D.N.Y. 2010).

¹⁴⁴ *See, e.g., Castaneda v. Burger King Corp.*, 259 F.R.D. 194, 197 (N.D. Cal. 2009) (holding that a request for pre-renovation measurements of a restaurant that was already under construction satisfied the requirement of undue hardship); *Fisher v. Kohl's Dept. Stores, Inc.*, No. 2:11-CV-3396 JAM GGH, 2012 WL 2377200, at *6 (E.D. Cal. June 22, 2012) (finding undue hardship in securing the contents of an incident report when both the sole witness and the plaintiff herself had no memory of the incident); *But see, Carr v. C.R. Bard, Inc.*, 297 F.R.D. 328, 334 (N.D. Ohio 2014) (holding that the plaintiff did not face undue hardship because the replication of a report was within the plaintiff's means); *In re Experian Data Breach Litig.*, No. SACV1501592AGDFMX, 2017 WL 4325583, at *3 (C.D. Cal. May 18, 2017) ("A showing of expense or inconvenience to Plaintiffs in hiring an expert to perform the same analysis isn't sufficient to overcome the protection of the work-product doctrine."); *Martin v. Bally's Park Place Hotel & Casino*, 983 F.2d 1252, 1263 (3d Cir. 1993) (finding a machine test enjoyed work-product protection

There are functional barriers that many practitioners face in accessing data necessary to build predictive models. Information may be limited by jurisdiction, by practice area, or by what metadata is available for analysis. Furthermore, empirical data on settlements is usually confidential.¹⁴⁵ Despite PACER’s vision to centralize federal court records, its implementation has been inconsistent and unreliable.¹⁴⁶ For these reasons, tools are understandably rolled out piecemeal, according to what data can be collected.¹⁴⁷ Courts typically also require that the evidence in question be essential or crucial to the requesting party’s case.¹⁴⁸ This requires that the information the party seeks is “an essential element” in its case, or of “great probative value on contested issues,” that cannot be obtained elsewhere.¹⁴⁹ “Need” is typically separate from hardship and linked to the importance of the information in prosecuting the case and making evidentiary showings.¹⁵⁰ The outputs of an outcome-prediction tool might aid a party’s prosecution of a litigation by, for instance, informing its settlement calculus, but it is not evidence to be addressed at trial.¹⁵¹ Thus, outcome predictions wouldn’t meet this standard of being “necessary” or relevant to proving or disproving the claims in the case.

C. Document Generation

Another type of burgeoning technology that may be the subject of discovery requests is software that generates legal documents. These tools draft documents from simple pleadings and answers to more complicated

because the party seeking to compel its production had the “technical capability” to replicate the analysis and the “resources necessary” to do so would not be “prohibitive”); *S.E.C. v. Thrasher*, No. 92 CIV. 6987 (JFK), 1995 WL 46681, at *10 (S.D.N.Y. Feb. 7, 1995) (holding the notes must be produced by comparing the cost of reproduction to the cost of securing the materials from the opposing party); *Burrow v. Forjas Taurus S.A.*, 334 F. Supp. 3d 1222, 1231-32 (S.D. Fla. 2018) (pointing to the moving party’s relative lack of knowledge of the steps required to reproduce the requested analysis and needed “facilities, equipment, tech, staffing, and expertise” in compelling production).

¹⁴⁵ Osbeck, *supra* note 112, at 62.

¹⁴⁶ Peter W. Martin, *District Court Opinions That Remain Hidden Despite a Long-Standing Congressional Mandate of Transparency – The Result of Judicial Autonomy and Systemic Indifference*, 110 L. LIBR. J. 305, 330 (2018).

¹⁴⁷ Artificiallawyer, *Bloomberg Law to Offer Lawyer-Client Representation Analysis*, ARTIFICIAL L., <https://www.artificiallawyer.com/2018/10/24/bloomberg-law-to-offer-lawyer-client-representation-analysis/> [https://perma.cc/5F7K-WKDV].

¹⁴⁸ See *Fletcher v. Union Pac. R.R. Co.*, 194 F.R.D. 666 (S.D. Cal. 2000)

¹⁴⁹ *Id.* at 672; *Nat’l Cong. for Puerto Rican Rts. v. City of New York*, 194 F.R.D. 105, 110 (S.D.N.Y. 2000).

¹⁵⁰ See, e.g., *Local 703 v. Regions Fin. Corp.*, 2012 WL 13027572, at *4 (S.D. Cal. Sep. 11, 2012) (finding a lack of “substantial need for the particular information sought, beyond [the] need to reduce . . . investigative costs by riding on [the opposing party’s] coattails”).

¹⁵¹ Engstrom, *supra* note 3, at n.312.

papers, such as discovery requests, motions, and even simple briefs.¹⁵² It is true that legal forms and templates are nothing new to the industry, but machine intelligence is revolutionizing their use.¹⁵³ Many of these programs allow a user to create a “coded” contract by uploading and coding a preexisting document with an accompanying questionnaire that can be used to quickly draft similar documents.¹⁵⁴

With current computer generation capabilities, non-lawyers can create the first draft of relatively boiler-plate documents before sending them along to an experienced associate who polishes them into finished products.¹⁵⁵ As the technology advances, these kinds of programs will be better able to provide drafts of briefs and memos, as well as connect to legal research programs, which will provide data for the writing program.¹⁵⁶

Relatedly, many technological tools can create, store, analyze, and monitor performance of contracts.¹⁵⁷ Legal departments of businesses are the prime users of such tools. These programs will increasingly be writing and monitoring the performance of contracts that govern transactions and become the subject of disputes. As adoption of these tools increases, the processes by which they operate may become of great interest to adverse parties. An adverse party may want this information to establish an element of the claim, such as the intent or expectations of the contracting parties. Importantly, this type of discovery would fall into the merit-directed discovery bucket, rather than the process-directed discovery bucket, because it goes to the underlying facts of the disputed document in the litigation.

Should the details surrounding these document generation tools be protected? Courts should make this determination by looking at the nature of the request and considering (1) the relevancy of the requested material; (2) the risk of exposing attorney mental impressions if disclosed; and (3) whether the attorney judgment involved was exercised in anticipation of litigation.

With respect to documents such as briefs or memos that are made in the context of litigation, those are unlikely to be relevant to the merits of the dispute and to assessing the quality of a production. However, if a contract or other document generated by a machine is heavily involved in a dispute,

¹⁵² Engstrom, *supra* note 3, at 1012.

¹⁵³ McGinnis, *supra* note 4, at 3050.

¹⁵⁴ Kathryn D. Betts & Kyle R. Jaep, *The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life into a Decades-Old Promise*, 15 DUKE L. & TECH. REV. 216, 219 (2017).

¹⁵⁵ *Id.*

¹⁵⁶ McGinnis, *supra* note 4, at 3052.

¹⁵⁷ Engstrom, *supra* note 3, at 1012. For a deeper explanation of predictive analytic technologies, see Betts, *supra* note 154.

then details surrounding its production or monitoring system may fall under Rule 26(b)'s scope of relevancy. We then think about the core of the work-product doctrine: attorney labor and mental impressions in anticipation of litigation. With document generators and monitors, the work is in developing the software, which require that humans train their models and regularly maintain them to refine their performance over time.¹⁵⁸ The lawyers and software developers must first collect all the best versions of the relevant document to create the template, then build the questionnaire in order to fill the template. This takes time and resources, such as access to a large enough document bank.

Once the software has produced a document draft or given a monitoring report, there is still work left for the lawyer. Admittedly, this part will shrink as the technology advances, but today, the lawyer still has significant strategizing or editing to do before he or she can take the next step. The machine will just produce the first draft; it will require follow up edits from one or multiple attorneys before it can be finalized. In these cases, the machine is analogous to the lower-level associate who makes the first draft and is just one step in a long line of edits. Thus, the machine outputs of document software are just the first step, which is followed by the attorney working on the document and leaving his or her mental impressions.

Tools used for contract generation, which do not perform their function in anticipation of litigation, would not be protected by work-product. In fact, they are utilized to create a better document, which typically means one that will avoid conflicts and litigation. Moreover, the details of these contract generators may go directly to proving or disproving elements of the claim being tried; they might be comparable to a company's procurement manual or guidelines. Thus, they will rightfully be discoverable in many situations. Tools used to generate documents in the midst of litigation, such as brief generators, would fall under the "in anticipation of litigation" umbrella. However, production of the details behind brief generators should not be compelled because they do not go to proving or disproving the underlying facts of the dispute, nor do they go to the sufficiency of a discovery production, such as a search term selection.

V. CONCLUSION

There are a complex set of factors to consider when assessing the discovery of legal technologies. For the reasons discussed above, judges

¹⁵⁸ Desai, *supra* note 102, at 28 (machine learning models are not just turned loose on data, rather, programmers make numerous decisions about how to partition data, which model types and data features to choose, and how much to tune the model); Katz, *supra* note 71, at 953 (the goal is to optimally update the model continuously as time ticks forward).

should exercise their discretionary power to decide whether to compel discovery upon motion from one of the parties in light of several considerations, including (1) the subject of the production and whether it is relevant to the underlying claims in the dispute or to assessing the adequacy of the adversary's production; (2) the risk of exposing attorney mental impressions; (3) the lifecycle of the development and usage of the technology, the point(s) at which attorney judgment is involved, and whether it was in anticipation of litigation.

TAR details related to the training and development of a program, including seed sets, should be discoverable only to the extent they are necessary to verify the validity of a discovery production in the litigation. TAR details related to the inputs for the specific relevant litigation, including search terms, should always be made available to the opposing party with any material claim of questionable production. With respect to predictive analytic programs, courts should not compel disclosure of their inputs or outputs because they will generally not be relevant to the merits of the case or the adequacy of a production. Regarding document generation, those that are assembled during litigation, such as briefs, should not be discoverable as they are irrelevant. Contract generators may be exposed to discovery because they are not protected by work-product doctrine as they aren't utilized in anticipation of litigation. As technologies continue to develop and the landscape evolves, litigants and judges will need to continue to reevaluate the application of old rules to new tools.