ARTICLE

LEGAL TECH, CIVIL PROCEDURE, AND THE FUTURE OF ADVERSARIALISM

DAVID FREEMAN ENGSTROM[†] & JONAH B. GELBACH[†]

"Legal tech" is transforming litigation and law practice, and its steady advance has tapped a rich vein of anxiety about the future of the legal profession. Much of the resulting debate narrowly focuses on what legal tech portends for the professional authority, and profitability, of lawyers. It is also profoundly futurist, full of references to "robolawyers" and "robojudges." Lost in this rush to foretell the future of lawyers and their robotic replacements is what should be an equally important, and also more immediate, concern: What effect will legal tech's continued advance have on core features of our civil justice system and, in particular, the procedural rules that structure it? Tackling that question, this Article seeks to enrich—and, in places, reorient—the budding debate about legal tech's implications for law and litigation by zeroing in on the near- to medium-term, not out at a distant, hazy horizon. It does so via three case studies, each one exploring how specific legal tech tools (e-discovery tools, outcome-prediction tools, and tools that perform advanced legal analytics) might alter litigation for good and ill by shifting the distribution of costs and information within the system. Each case study then traces how a concrete set of civil procedure rules—from Twombly/Iqbal's pleading standard and the work product doctrine to rules and doctrines that govern forum-shopping—can, or should, adapt in response. When these assorted dynamics are lined up and viewed together, it is not a stretch to suggest that legal tech will remake the adversarial system, not by replacing lawyers and judges with robots, but rather by unsettling, and even resetting, several of the system's procedural cornerstones. The challenge for courts—and, in time, for rulemakers and legislators—

(1001)

will be how best to adapt a digitized litigation system using civil procedure rules built for a very different, analog era. This Article aims to jumpstart thinking about that process by identifying the principal ways that legal tech will reshape "our adversarialism" and mapping a reform and research agenda going forward.

INTRODUCTION

"Legal tech," most agree, is transforming litigation and law practice, and its steady advance has tapped a rich vein of anxiety about the future of the

legal profession.[1] Is law like a driverless car, or is it irreducibly complex and grounded in dynamic human judgment? How to square online dispute resolution and automated legal advice with rules governing unauthorized practice of law? Can BigLaw survive? Much of this has a profession-centered and even defensive quality in its narrow focus on what legal tech portends for the professional authority and profitability of lawyers. Much of it is also profoundly futurist—full of prophecies of "robolawyers,"[2] "robojudges,"[3] or

---

[1] *See, e.g.*, BENJAMIN H. BARTON, GLASS HALF FULL: THE DECLINE AND REBIRTH OF THE LEGAL PROFESSION (2015); RICHARD SUSSKIND, TOMORROW'S LAWYERS: AN INTRODUCTION TO YOUR FUTURE (2013); RICHARD SUSSKIND & DANIEL SUSSKIND, THE FUTURE OF THE PROFESSIONS: HOW TECHNOLOGY WILL TRANSFORM THE WORK OF HUMAN EXPERTS (2015); 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 (2013); John O. McGinnis & Russell 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 (2014); Dana Remus & Frank Levy, *Can Robots Be Lawyers?: Computers, Lawyers, and the Practice of Law*, 30 GEO. J. LEGAL ETHICS 501 (2017); Tanina Rostain, *Robots versus Lawyers: A User-Centered Approach*, 30 GEO. J. LEGAL ETHICS 559 (2017); Eric L. Talley, *Is the Future of Law a Driverless Car?: Assessing How the Data Analytics Revolution Will Transform Legal Practice*, 174 J. INST. & THEORETICAL ECON. 183 (2018); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117 (2014); Bruce H. Kobayashi and Larry E. Ribstein, *Law's Information Revolution*, 53 ARIZ. L. REV. 1169 (2011); William D. Henderson, *A Blueprint for Change*, 40 PEPP. L. REV. 461 (2013). For treatments in the popular and lawyer-trade presses, see Mark A. Cohen, *'Legal Innovation' Is Not an Oxymoron—It's Further Along Than You Think*, FORBES (Mar. 14, 2017, 8:59 AM), https://www.forbes.com/sites/markcohen1/2017/03/14/legal-innovation-is-not-an-oxymoron-its-farther-along-than-you-think [https://perma.cc/85TT-XZY4]; Jason Koebler, *Rise of the Robolawyers*, ATLANTIC, Apr. 2017, https://www.theatlantic.com/magazine/archive/2017/04/rise-of-the-robolawyers/517794 [https://perma.cc/QN8P-NLK4]; Steve Lohr, *AI Is Doing Legal Work. But It Won't Replace Lawyers, Yet.*, N.Y. TIMES (Mar. 19, 2017), https://www.nytimes.com/2017/03/19/technology/lawyers-artificial-intelligence.html [https://perma.cc/Z7ZX-V3WB]; John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. TIMES (Mar. 4, 2011), https://www.nytimes.com/2011/03/05/science/05legal.html [https://perma.cc/JD4T-JK3B]; Debra Cassens Weiss, *Will Technology Create a Lawyer 'Jobs-Pocalypse'? Doomsayers Overstate Impact, Study Says*, A.B.A. J. (Jan. 5, 2016, 6:45 AM), http://www.abajournal.com/news/article/does_technology_presage_a_lawyer_jobs_pocalypse_naysayers_overstate_impact [https://perma.cc/36AT-R7L3]; John G. Browning, *Will Robot Lawyers Take Our Jobs?*, D CEO MAG., Mar. 2019, https://www.dmagazine.com/publications/d-ceo/2019/march/will-robot-lawyers-take-our-jobs [https://perma.cc/W6EZ-A4Q9].

[2] *See, e.g.*, Asa Fitch, *Would You Trust a Lawyer Bot with Your Legal Needs?*, WALL. ST. J. (Aug. 10, 2020, 10:00 AM), https://www.wsj.com/articles/would-you-trust-a-lawyer-bot-with-your-legal-needs-11597068042 [https://perma.cc/72LX-B72J]; Gary Marchant & Josh Covey, *Robo-Lawyers: Your New Best Friend or Your Worst Nightmare?*, LITIGATION, Fall 2018, at 27; Koebler, *supra* note 1. Not everyone is so confident. *See* Milan Markovic, *Rise of the Robot Lawyers?*, 61 ARIZ. L. REV. 325, 325 (2019) (challenging the notion that lawyers will be displaced by artificial intelligence).

[3] The literature that predicts or otherwise assumes a future populated by robojudges is growing fast. *See, e.g.*, Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135 (2019) (predicting a future with robot judges); Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611 (2020) (same); Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1 (2019) (examining the eventual flaws in machine-led law); Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, 109 J. CRIM. L. & CRIMINOLOGY 137 (2019) (examining the merits of AI-run legal decisionmaking); Rebecca Crootof, *"Cyborg Justice" and the Risk of Technological-Legal Lock-In*, 119 COLUM. L. REV. F. 233 (2019) (discussing the uptick in AI adjudication); Eric Niiler, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, WIRED (Mar. 25, 2019, 7:00 AM), https://www.wired.com/story/can-ai-be-fair-

even an eventual state of "legal singularity,"[4] when machines can perfectly predict the outcomes of cases before they are filed.

Lost in this rush to foretell the future of lawyers and their robotic replacements is what should be an equally or even more important concern: what effect will legal tech's continued advance have on core features of our civil justice system and, in particular, the procedural rules that structure it? And how, in turn, can or should those rules be adapted to further the ends of justice? This Article seeks to enrich—and, in places, reorient—budding debate about what many see as a coming revolution in legal tech. Simply put, if law and the legal profession will look different ten or fifteen years from now, then civil procedure and the inner workings and structure of the adversarial system will look different as well. Indeed, though virtually unmentioned in a lively but high-altitude new literature on legal tech's potential implications, it is the rules of civil procedure and related doctrines that will serve as the front-line regulators of the new legal tech tools and critically shape their evolution in the near-to medium-term. As a result, judges, rulemakers, and legislators should begin to think about whether, and if so how, to adapt civil procedure to new litigation realities as legal tech continues its move to the center of the civil justice system.

We aim to spark concrete thinking about this mediating role for civil procedure by focusing on the near future—not out at a hazy horizon dotted with robojudges and robolawyers—and then asking how legal tech will change litigation and, in turn, how procedure can or should adapt in response.[5] The core of our argument proceeds from the premise that legal tech's proliferation

judge-court-estonia-thinks-so [https://perma.cc/ES5C-G9W6] (discussing how Estonia has embraced AI in court proceedings); Christopher Markou, *Are We Ready for Robot Judges?*, DISCOVER MAGAZINE (May 15, 2017, 8:00 PM), https://www.discovermagazine.com/technology/are-we-ready-for-robot-judges [https://perma.cc/E7AM-PWCH] (examining the increased interest in AI judges); AirTalk Podcast, *Can a Robot Make a Fair Verdict?*, 89.3 KPCC, https://www.scpr.org/programs/airtalk/2019/04/01/64335/can-a-robot-judge-make-a-fair-verdict [https://perma.cc/855F-VMWH] (discussing AI capabilities in adjudication); Victor Tangermann, *Estonia is Building A "Robot Judge" to Help Clear Legal Backlog*, FUTURISM (Mar. 25, 2019), https://futurism.com/the-byte/estonia-robot-judge [https://perma.cc/P3YV-TJ4U] (examining Estonia's new use of AI for adjudication); H.W.R. (Henriëtte) Nakad-Weststrate, Ton Jongbloed, H.J. (Jaap) van den Herik, Abdel-Badeeh M. Salem, *Digitally Produced Judgements in Modern Court Proceedings*, 6 INT'L J. DIGIT. SOC. 1102 (2015) (investigating the use of AI verdicts in arbitration); *see also* Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 242 (2019) ("[T]he prospect of 'robot judges' suddenly seems plausible—even imminent.").

    4 *See* Benjamin Alarie, *The Path of the Law: Towards Legal Singularity*, 66 U. TORONTO L.J. 443, 443 (2016).

    5 A growing body of opinion holds that a rigorous focus on near- to mid-term development and issues is both necessary and under-supplied. *See, e.g.*, Edward Parsons, Alona Fyshe & Dan Lizotte, *Artificial Intelligence's Societal Impacts, Governance, and Ethics*, UCLA: THE PROGRAM ON UNDERSTANDING LAW, SCIENCE, AND EVIDENCE (PULSE) (2019), https://escholarship.org/uc/item/2gp9314r [https://perma.cc/5X7Y-TEWZ] (noting a "bi-modal" distribution of inquiry, with some attracted to "speculative" thinking about "endpoint, singularity-related issues," and others more disciplinarily inclined toward "current concerns and historical precedents," leaving a "disturbingly empty [and] large middle ground of impacts and challenges lying between these endpoints").

is likely to alter two foundational aspects of any litigation system: the distribution of litigation costs and the distribution of information. In a nutshell, there is good reason to believe that the concern about high and asymmetric litigation costs that has fueled several decades' worth of litigation reforms will progressively fade as new and powerful e-discovery tools propagate. By contrast, it is plausible that increasing uptake of legal tech tools, including e-discovery tools but also tools that perform legal research and analytics and predict case outcomes, will worryingly widen information asymmetries within the system, between judges and litigants, and also between litigants and litigants—particularly litigation's "haves" as against its "have nots."

Isolating legal tech's effects on these deep dimensions of the system provides needed analytic traction and grounds a set of concrete judgments about how an array of civil procedure rules and doctrines—among them the plausibility pleading standard set forth in *Twombly* and *Iqbal*, the bundle of procedural rules, doctrines, and statutes concerned with forum-shopping, and the work product doctrine—can, or should, adjust in response. When these assorted dynamics are lined up and viewed together, it is not a stretch to say that legal tech will, in time, remake the adversarial system, not by replacing lawyers and judges with robots, but rather by unsettling, and even resetting, several of its procedural cornerstones.

These are big claims, and they demand both a technical grasp of the legal tech toolkit and command of contemporary civil procedure. Given these complexities, we build our argument deliberately, in three steps.

Part I offers a full and quasi-technical canvass of where legal tech currently is and where it is likely to go in the near- to medium-term as natural language processing (NLP) and other machine learning techniques that power the most consequential legal tech tools continue to improve. In so doing, we strike a skeptical note and also go about our labors with a heavy dose of humility. As with any emergent technology, legal tech is a fast-moving field, and any effort to capture its many facets risks becoming antiquated almost as soon as the ink dries. We manage this contingency by surveying the legal tech landscape in three pieces. Section I.A reviews legal tech's flavors and offers some ways to slice and dice them. Section I.B turns to legal tech's technical trajectory. It shows that the frontier is quickly moving beyond e-discovery and digital referencing tools (think Westlaw or Lexis) to tools that automatically gather legal materials, predict case outcomes, and even draft legal documents. However, there are also legitimate questions about how far and how quickly legal tech can advance. Just how much progress can be made on outcome prediction tools given pervasive confidential settlements and the resulting lack of well-labeled data, or the current technical limits of natural language processing (NLP) in extracting and analyzing legal argumentation? Section I.C summarizes some key implications of legal tech—for the legal profession, for the distribution of power within the legal system, and for law itself—as sketched in an emerging

academic literature that, while highly abstract, has begun to stake out the poles of a rich debate. A thorough survey of the legal tech landscape provides the raw material for the more focused case studies of procedure to come.

Armed with Part I's extended account of legal tech's pathways of innovation and diffusion, we turn in Part II to offering three concrete cuts at how legal tech's advance will reshape American litigation and how procedural rules might mediate those effects.

Section II.A starts on familiar ground: e-discovery and, more specifically, the "technology-assisted review" (TAR) and "predictive coding" tools that are quickly becoming a fixture of complex litigation practice. Our core claim is that, contrary to the views of some, civil litigation may well see a steady decline in overall discovery costs and, by extension, a narrowing of the litigation cost asymmetries that have motivated decades of litigation reforms, from the Civil Justice Reform Act of 1990 to the 2015 amendments reshuffling Rule 26's proportionality constraint. In a lower-friction world, we predict, battles over proportionality would largely abate or become peripheral. Narrowing litigation cost asymmetries may also alter, or at least destabilize, the normative foundation of a very different and controversial part of civil procedure: the plausibility pleading doctrine set forth in *Twombly* and *Iqbal*. That doctrine sits at the intersection of two competing concerns: litigation cost asymmetries, with attendant concerns about undue settlement leverage and the conversion of low- or even negative-value cases into positive-dollar settlements, and information asymmetries in cases where only discovery can dislodge privately held information about wrongdoing. By systematically narrowing litigation cost asymmetries, TAR could undermine the positive foundation of the new plausibility pleading regime.

Section II.B turns to legal tech tools that predict case outcomes. An obvious concern is that continued advances in outcome prediction tools will foster forum shopping, placing pressure on the rules, statutes, and doctrines—venue, removal, *Erie* doctrine—that seek to limit or shape its pursuit. Here we sound a more skeptical note about legal tech's implications for civil procedure. Current procedural rules and doctrines touching upon forum shopping strike a permissive pose, and so an initial question is whether successful deployment of predictive analytics *should* change that pose. Further grounds for skepticism are the technical and practical limits of outcome prediction tools, which may not "work" well enough to meaningfully increase forum-shopping in the first place. Even if the technical hurdles can be leapt, two other major obstacles stand in legal tech's way. First is the huge cost of assembling enough docket and document data to obtain sufficiently large and representative samples for contemporary prediction methods. Second are a pair of endogeneity problems that raise profound questions about either the initial or subsequent usefulness of even predictions based on large samples of data. Still, we think it useful to ask: if predictive analytics did "work," and if machine-aided

forum shopping falls into disfavor, what follows? Here lie some of the most bracing procedural possibilities. Effective outcome-prediction tools and supercharged forum-shopping might steadily widen asymmetries in the quantity and quality of information available to litigants and judges. This might warrant changes in the treatment of forum shopping motives, in the discoverability of work product, or both. It would also raise questions about whether judges making choice-of-forum determinations, or deciding motions to dismiss and for summary judgment, should be empowered to order parties to disclose their machine outputs or perhaps should even be equipped with the same prediction tools litigants are using. Either scenario would press on the bounds of current conceptions of "managerial judging" and the proper allocation of authority between judge and jury.

Section II.C asks a key question that looms in the background of the other case studies and, indeed, all of legal tech: how might the work product doctrine need to change to accommodate a world in which a non-trivial amount of lawyering, including not just discovery and outcome prediction, but also legal research, brief writing, and strategic litigation judgment, takes the form of machine-generated outputs? The fount of the work product rule, *Hickman v. Taylor*, famously brackets distributive concerns—i.e., the fact that some parties can afford better counsel than others—and instead protects against "wits borrowed from the adversary," as Justice Jackson put it, so that parties, and the system, can capture the benefit of good lawyering. In so doing, the work product rule secures the conditions necessary for a well-functioning adversarial system by ensuring returns on, and thus investment in, legal talent. But as legal tech tools grow more powerful, and if the "haves" have them and the "have nots" do not, legal tech could well shift the normative ground out from under a cornerstone of the American procedural system. In this new machine-driven world, should we, to invoke Justice Jackson's turn of phrase in *Hickman*, protect against "borrowed bits" the same way we protect against "borrowed wits"?

Part III steps back and draws out some connections across the case studies. In particular, we show how legal tech's continued advance will place civil procedure in a new and uncharted posture. In particular, judges—and, in time, rulemakers and legislators—will come to preside over what amounts to a shadow innovation policy because their procedural choices will shape the terms of legal tech's use, its value to litigants, and its market for production. Just as important, legal tech's advance will compel judges and policymakers to make explicit or implicit judgments about the optimal balance of adversary as against judicial control of civil proceedings. In making those judgments, they will shape the future of American adversarialism.

Before launching, some caveats: First, we bracket the criminal context entirely, despite a rich and growing literature on the use of predictive analytics

to support decisions about bail and sentencing.[6] Second, we make no claims to comprehensiveness, nor is ours a case study approach in the rigorous comparative sense of making causal judgments about legal tech's effect on procedure or vice versa. Rather, our aim is to map legal tech's conceptual landscape and, by identifying a set of key procedural questions it implicates, chart further productive lines of inquiry. Last, we seek to be both far-thinking and concrete, thus achieving a salutary, middle-level of abstraction that is grounded in actual, not hypothetical, legal tech tools, as mediated by existing, not hypothetical, procedural rules. In other words, we aim to cut through the "AI fever" that infects the literature on legal tech, and on AI and law more broadly, without losing generality or zing. This is not to say legal tech lacks implications for the civil justice system beyond civil procedure. Will legal tech further vanish the vanishing trial, blunt incentives for private litigants to conduct socially valuable discovery, or stunt the dynamic evolution of legal norms? Throughout the Article, we address these and other wider-aperture questions in passing. However, our focus remains how procedure can, or should, mediate the legal tech revolution over the near- to medium-term.

I. THE LEGAL TECH LANDSCAPE

In *Law's Empire*, Ronald Dworkin builds his influential theory of legal interpretation around a mythical uber-judge, Hercules, with a superhuman capacity to read and understand every available scrap of legal material and thus reach a unique right answer in every case.[7] Dworkin's project was to critique the legal positivism of H.L.A. Hart, and so it sits far away, intellectually speaking, from the world of legal tech. But Dworkin's Hercules has recently taken on renewed relevance as legal tech's advance has allowed us to glimpse a world in which machines, not just mythical judges, can unerringly adjudicate cases or even predict a case's outcome before it is filed. As Michael Livermore and Daniel Rockmore recently put it, judges, lawyers, and much of the legal system as we know it may someday soon be replaced by "blinking computerized Herculi" that sit in "server farms rather than law offices."[8]

This image of server farms replacing courthouses is an emotive one, and a cottage industry of mostly academic commentators has seized on it and set about imagining futuristic endpoints—an event horizon of sorts, where law

---

6 For two excellent entries in a vast and growing literature examining the growing use of algorithmic tools, see Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043 (2019) and Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019).

7 *See generally* RONALD DWORKIN, LAW'S EMPIRE (1986).

8 *See* Michael A. Livermore & Daniel N. Rockmore, *Introduction: From Analogue to Digital Legal Scholarship*, *in* LAW AS DATA: COMPUTATION, TEXT, & THE FUTURE OF LEGAL ANALYSIS xii, xiv (Michael A. Livermore & Daniel Rockmore eds., 2019) [hereinafter LAW AS DATA].

and technology meet. Blinking Herculi, it is said, will in turn bring a state of "legal singularity," when all legal outcomes are perfectly predictable ex ante, and all uncertainty is banished from the system.[9] Law itself will be steadily transformed into a "catalog of precisely tailored laws" or "microdirectives"[10] made up of "up-to-the-second" and "individualized"[11] rules that adjust in real-time—for instance, an individualized speed limit for a given driver with a given amount of experience operating in specific driving conditions—and are enforced via automatic penalties.[12] As this new and "seamless legal order" settles into place, there is no longer any need, or any room, for lawyering, adjudication, judges, or judicial discretion. Law becomes "self-driving."[13]

But just how likely are we to get there, and, assuming we make it at all, how soon? More importantly, what can we expect in the meantime? This Part addresses these questions. In so doing, we lower our gaze to a useful middle distance—our eyes neither inside the boat nor drifting out to a distant, Herculi-blinking horizon—and provide a systematic accounting of what we currently know, and don't know, about the state of legal tech. We address legal tech's current range of applications (Section I.A); its trajectory, as shaped in particular by the technological possibilities and limits of text-based analytics (Section I.B); and its implications for the legal profession, the legal system, and law itself (Section I.C). The resulting composite portrait provides the raw materials necessary for Part II's exploration of some concrete ways the rules of civil procedure will mediate legal tech's incorporation into the adversarial system in the near- to medium-term.

A. *Flavors of Legal Tech*

An initial task is to survey legal tech's sprawling landscape. In what ways, and toward what ends, are legal tech tools being deployed within the civil justice system?

A small but growing literature sizes up the legal tech field and offers some ways to slice and dice its component parts.[14] One approach honors legal tech's

---

9 *See* Alarie, *supra* note 4, at 445.

10 *See* Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401, 1402-03 (2017).

11 Brian Sheppard, *Warming Up to Inscrutability: How Technology Could Challenge Our Concept of Law*, 68 U. TORONTO L.J. 36, 38 (Supp. 1, 2018); *see also* Casey & Niblett, *supra* note 10, at 1404-06 (explaining microdirectives); Anthony J. Casey & Anthony Niblett, *Self-Driving Contracts*, 43 J. CORP. L. 1, 7-10 (2017) (providing examples of microdirectives).

12 *See* Sheppard, *supra* note 11, at 40 (offering the example of regulating traffic to reduce congestion); Casey & Niblett, *supra* note 10, at 1404; Casey & Niblett, *supra* note 11, at 7-10 (same).

13 *See* Anthony J. Casey & Anthony Niblett, *Self-Driving Laws*, 66 U. TORONTO L.J. 429, 430 (2016) ("The law will become, for all intents and purposes, 'self-driving.'")

14 *See, e.g.*, KEVIN D. ASHLEY, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS: NEW TOOLS FOR LAW PRACTICE IN THE DIGITAL AGE (2017) (explaining how computational processes

entrepreneurial tilt and focuses on the sales channel, categorizing tools based on their end users (e.g., lawyers, clients/parties, businesses).[15] Among its virtues, this approach separates out tools that substitute for legal representation (e.g., online legal advice tools) from those that remain within lawyers' locus of control (e.g., e-discovery tools). Another approach could focus on the task performed: legal research, document management and creation, and document- and case-level analytics, among others. Still another approach could focus on the point in litigation time, beginning at the front-end of a case at which a tool is used and progressing forward: lawyer-client matching, legal research and analysis, discovery, the drafting of pleadings and documents, and trial. A final approach could categorize legal tech tools based on subject area. This approach highlights proliferating domain-specific tools, particularly in the contracts area, but also patents (e.g., tools that value patents and patent portfolios), divorce (the area closest to fully automated generation of legal documents), torts (where case valuation tools are most regularly in use), and tax (where legal analytics and prediction tools appear most advanced), with the rest allocated to a residual, "general" category.

The reality is that none of these approaches will be mutually exclusive and collectively exhaustive. Instead, Table 1 offers a mash-up of approaches in an effort to capture, at a glance, the main contours of the legal tech terrain. The result is nine categories of tools.

---

will change the practice of law by breaking down the multitude of areas of expected change); Benjamin Alarie, Anthony Niblett & Albert H. Yoon, *How Artificial Intelligence Will Affect the Practice of Law*, 68 U. TORONTO L.J. 106 (Supp. 1, 2018); Daniel Ben-Ari, Yael Frish, Adam Lazovski, Uriel Eldan & Dov Greenbaum, *Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment*, 23 RICH. J.L. & TECH., no. 2, art. 3, 2017, https://jolt.richmond.edu/files/2017/03/Greenbaum-Final-2.pdf [https://perma.cc/NXY3-7GZ2] (dividing the field into two components: strong and weak); Kathryn D. Betts & Kyle Jaep, *The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life Into a Decades-Old Promise*, 15 DUKE L. & TECH. REV. 216 (2017) (examining technological innovation in contract drafting); Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, 17 RICH. J.L. & TECH., no. 3, art. 11, 2011, http://jolt.richmond.edu/jolt-archive/v17i3/article11.pdf [https://perma.cc/7CEE-S7FN]; David Lat & Brian Dalton, *How Artificial Intelligence Is Transforming Legal Research*, ABOVE THE L. (July 16, 2018, 9:35 AM), https://abovethelaw.com/2018/07/how-artificial-intelligence-is-transforming-legal-research [https://perma.cc/VL27-2VWZ]. For a view of the landscape from within the industry, see THOMSONREUTERS, ALTERNATIVE LEGAL SERVICE PROVIDERS 2019, https://legal.thomsonreuters.com/en/insights/reports/alternative-legal-service-provider-study-2019.

   15 *See* Daniel W. Linna Jr., *What We Know and Need to Know About Legal Startups*, 67 S.C. L. REV. 389, 402-03 (2016).

Table 1: "Legal Tech" at a Glance

Category	End User	Description	Litigation Time	Examples
Lawyer Marketplace and Matching	lawyers, litigants, businesses	Tools that match lawyers with clients or facilitate a would-be client's evaluation and choice of potential counsel. This category also includes tools that assist lawyers with business and client development via assessments of a current or potential client's past legal entanglements or present legal exposure.	pre-filing	Avvo, Atticus, Priori Legal
Legal (Re)Search	lawyers	Tools that help lawyers locate and gather relevant raw materials (caselaw, statutes, regulations).	pre-filing, throughout litigation	CaseText, Ross Intelligence
Outcome Prediction	lawyers, businesses (including litigation financiers)	Tools that predict case outcomes. Predictions can be jurisdiction- or judge-specific and are used to compare forums and assess case quality at intake, filing (i.e., forum-shopping), or, once litigation is underway, to inform the strategic litigation and settlement calculus by predicting likely damages, how the assigned judge will rule, and the likely case stage and timeframe for resolution.	pre-filing, throughout litigation	Colossus, Ravel, Lex Machina, Gavelytics, Blue J Legal
Legal Analytics	lawyers, businesses	Tools that perform analytic tasks other than legal search and outcome prediction, including citation mappings, judge-level analytics (e.g., tailoring arguments to a specific judge) and document-level analytics (e.g., brief evaluation).	throughout litigation	Ravel, FastCase, Gavelytics, Lex Machina
Discovery	lawyers	Tools that support or supplant the process of identifying relevant documents and tagging them for privilege.	discovery phase	Everlaw, Relativity, OpenText, Exterro, CS Disco

Table 1: "Legal Tech" at a Glance (cont.)

Category	End User	Description	Litigation Time	Examples
Document Assembly and Creation	lawyers, litigants	Tools that draft legal documents, from simple pleadings (answers) to more complicated pleadings and papers (discovery requests, motions, and even simple briefs).	pre-filing and throughout litigation	Legalmation, RocketLawyer, CaseText, Lawyaw
Practice Management	lawyers	Litigation conduct tools, including dashboards that manage client intake, organize key case facts and documents, and support billing or other administrative tasks.	throughout litigation, post-litigation	Needles, Clio, Legal Server
Contract Management and Analysis	lawyers, businesses	Tools that store, analyze, create, and monitor performance of contracts.	pre-litigation	Kira Systems, Ravn, eBrevia, LexCheck, UnitedLex, LawGeex, Ironclad, Knowable, Evisort
DIY Dispute Resolution, Online Legal Advice, Court Vendor Services	lawyers, businesses	Tools that facilitate extra-legal resolution of disputes; tools that provide automated (often online) legal advice or assist unrepresented litigants with legal proceedings.	pre-filing, throughout litigation	LegalZoom, RocketLawyer, Modria, Intraspexion, Nolo, Matterhorn, Houston.AI, FairShake, DoNotPay
Document Assembly and Creation	lawyers, litigants	Tools that draft legal documents, from simple pleadings (answers) to more complicated pleadings and papers (discovery requests, motions, and even simple briefs).	pre-filing and throughout litigation	Legalmation, RocketLawyer, CaseText, Lawyaw

A further task is to identity the subset of legal tech tools that are most likely to play a central role in the legal system going forward and, in particular, will press most strongly on its adversarial structure and procedural rules. This requires more than a laundry list of applications. We need to understand how existing legal tech tools intersect with the system, and we also need to look under their hood and understand their technical and operational details. Toward that end, consider three further ways to carve up the field.

First, legal tech tools vary based on whether they operate inside or outside the litigation system and, in turn, whether they implicate procedural rules as opposed to other rules or policies. E-discovery, legal research, legal analytics, and outcome prediction tools all operate squarely *within* litigation because they assist lawyers or litigants seeking judicial resolution of disputes. As Part II will argue, they thus press on, and may even reshape, a range of civil procedure rules. Other legal tech tools, however, largely operate *outside* the litigation system—and, indeed, may seek to supplant it.[16] Table 1's DIY dispute resolution systems plainly fit this mold. So might automated (and typically online) legal advice systems in light of the large mass of disputes that are currently resolved, with little or no court proceedings, via direct negotiation by injured parties with insurance companies, court-ordered ADR (e.g., mediation), or arbitration.[17] To be sure, these latter tools hold implications for the litigation system. They shrink its domain. And by fueling alternative modes of dispute resolution, they exert pressure on the litigation system to adapt and can thus shape its inner procedural workings. Still, these tools more directly implicate legal-ethical rules sounding in consumer protection, such as unauthorized practice of law and solicitation restraints, than procedure.[18]

---

16 For a recent overview of legal tools serving the unrepresented, see REBECCA L. SANDEFUR, LEGAL TECH FOR NON-LAWYERS: REPORT OF THE SURVEY OF US LEGAL TECHNOLOGIES (2019) (surveying the tools available to non-lawyers). *See also* J.J. Prescott, *Improving Access to Justice in State Courts with Platform Technology*, 70 VAND. L. REV. 1993 (2017) (investigating the impacts of new online dispute resolution (ODR) technologies); Amy J. Schmitz, *Expanding Access to Remedies Through E-Court Initiatives*, 67 BUFF. L. REV. 89, 101 (2019); Amy L. Schmitz, *Dangers of Digitizing Due Process*, (Univ. Mo. Sch. L., Working Paper No. 2020-01, 2020) (examining the possible downfalls of computerized law).

17 *See* NAT'L CTR. FOR STATE CTS. & STATE JUST. INST., THE LANDSCAPE OF CIVIL LITIGATION IN STATE COURTS v (2015), https://www.ncsc.org/~/media/Files/PDF/Research/CivilJusticeReport-2015.ashx [https://perma.cc/W4B4-YG3H] (finding most litigants with resources have "already abandoned the civil justice system" through contract or private ADR). Even smart contracting may qualify given that it may obviate the need for any adjudication at all. Kevin Werbach & Nicolas Cornell, *Contracts Ex Machina*, 67 DUKE L.J. 313, 339 (2017).

18 Other tools, particularly court vendor companies, have significant access to justice implications but fewer procedural implications. *See Courts & Judiciary*, THOMSON REUTERS, https://legal.thomsonreuters.com/en/solutions/government/courts-judiciary [https://perma.cc/3F7R-FMJJ] (offering court-oriented technologies); *Courts & Justice*, TYLER TECHS., https://www.tylertech.com/solutions/courts-public-safety/courts-justice [https://perma.cc/5VPV-WBY2] (offering technologies for courts and justice agencies). The exception is online dispute resolution (ODR) platforms, which raise a host of procedural issues, including "traffic rules" for moving litigants

Second, legal tech tools plainly differ in their technical sophistication and their degree of advance over analog legal practice. Many of the business development tools in Table 1's "Lawyer Marketplace and Matching" category may be little more than glorified docket monitoring.[19] Similarly, some tools falling into the "Legal (Re)search" category are merely more feature-rich versions of search platforms like Westlaw and Lexis that have long been part of the lawyer's workbench. These tools offer enhanced filtering capacities— e.g., by judge, or by procedural posture—or improved user interfaces but otherwise provide much the same basic service as incumbent tools.[20] Other tools, however, go well beyond lawyer-directed digital referencing by permitting a lawyer to drag and drop a complaint or brief and receive on-point cases (i.e., cases sharing facts, legal issues, and jurisdiction). Still other tools feel different in kind. After ingesting only the pleadings and papers to that point in a specialized tax or labor litigation, some advanced legal tech tools can generate a simple draft motion or brief or response to an agency's civil investigative demand at the touch of a button.[21] Table 1 attempts to capture this distinction across the "Legal (Re)search" and "Legal Analytics" categories, with the former skewing toward lawyer-controlled tools that return less digested baskets of legal materials—a kind of "hunting and gathering"—and the latter encompassing, and aspiring to, more advanced legal cognitions.[22]

---

into and out of ODR platforms, and also "information rules" that govern what information is provided to disputants on the platform about their options and prospects to nudge them toward settlement. *See* David Freeman Engstrom, *Digital Civil Procedure*, 169 U. PA. L. REV. (forthcoming 2021).

19 *See* Patrick Flanagan & Michelle H. Dewey, *Where Do We Go from Here? Transformation and Acceleration of Legal Analytics in Practice*, 35 GA. ST. U. L. REV. 1245, 1253 (2019).

20 An example is Judicata, which highlights its ability to filter based on appealing party, cause of action, court, and procedural posture. *See Introducing Clerk*, JUDICATA, https://www.judicata.com [https://perma.cc/7YCW-4PJJ] (noting "advanced filters"); *see also* ROSS INTELLIGENCE, https://www.rossintelligence.com [https://perma.cc/2KDX-VFGZ] (claiming to be "easier to use than Westlaw and LexisNexis" and providing a feature-rich search tool based on unique case facts and procedural posture).

21 *See* Susan Beck, *Inside ROSS: What Artificial Intelligence Means for Your Firm*, LAW.COM (Sept. 28, 2016), http://www.law.com/sites/almstaff/2016/09/28/inside-ross-what-artificial-intelligence-means-for-your-firm [https://perma.cc/W3P6-VHWT] (discussing ROSS's ability to draft motions); Rob Carty, *Computer-Written Legal Briefs Are Closer Than You Think*, ARTIFICIAL LAWYER (Apr. 11, 2019), https://www.artificiallawyer.com/2019/04/11/computer-written-legal-briefs-are-closer-than-you-think [https://perma.cc/FJ9Z-Y49L] (reviewing the current abilities of briefs written using A.I.).

22 *See, e.g.*, Faraz Dadgostari, Mauricio Guim, Peter A. Beling, Michael A. Livermore & Daniel N. Rockmore, *Law Search as Prediction*, A.I. & LAW (2020) (formulating a mathematical model with a more complex search process); Kevin D. Ashley, *Automatically Extracting Meaning from Legal Texts: Opportunities and Challenges*, 35 GA. ST. U. L. REV. 1117, 1133-35 (2019) (arguing that "QA" tools that allow a lawyer to ask questions in natural language are a key frontier of legal tech); Kingsley Martin, *Deconstructing Contracts: Contract Analytics and Contract Standards*, *in* DATA-DRIVEN LAW: DATA ANALYTICS AND THE NEW LEGAL SERVICES 37 (Ed Walters ed., 2018) (describing automated contracting); Albert H. Yoon, *The Post-Modern Lawyer: Technology and the Democratization of Legal Representation*, 66 U. TORONTO L.J. 456, 457 (2016) (noting that most promising tools focus "not

A further generalization regarding technical sophistication is that the most potentially game-changing legal tech tools perform prediction tasks and incorporate one or more elements of machine learning (ML). The first part of this—a focus on prediction—should not surprise. Litigation takes place in the "shadow of the law," as Mnookin and Kornhauser famously put it,[23] and much of lawyering involves making predictive judgments in that shadow.[24] Which cases are winners and which losers? Which documents are relevant, and which can be defensibly withheld on privilege grounds? And which legal arguments and precedents will this judge find most persuasive? Machine prediction tools aim to replicate these fundamentally predictive cognitions.

Nor should it surprise that the most promising legal tech tools deploy ML. For the uninitiated, machine learning is a family of algorithm-based techniques that use statistical models to "learn" from data in specific contexts rather than relying on more structured rules that an analyst programs directly.[25] Beyond this high-level commonality, however, ML methods are a varied lot, and the techniques that power legal tech are no exception. First, many legal tech tools use "supervised" ML methods that analyze a set of previously and typically human-labeled data inputs—referred to as "training data"—in order to draw predictive inferences about the labels humans would assign to new and unseen instances.[26] At least for the moment, fewer legal

on replacing lawyers' routinized tasks but, rather, on facilitating how they understand and analyze legal materials"). A good example is CaseText's CARA, which allows users to input a brief and receive back a list and summary of cases. *See* Valerie McConnell, *What is CARA A.I. and How Do I Use It?*, CASETEXT, https://help.casetext.com/en/articles/1971642-what-is-cara-a-i-and-how-do-i-use-it [https://perma.cc/94L6-73ER]. For an engaging history of the evolution of computerized legal search, see John O. McGinnis & Steven Wasick, *Law's Algorithm*, 66 FLA. L. REV. 991, 991 (2014).

   23 *See* Robert N. Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950, 950 (1979).

   24 *See* Mark K. Osbeck, *Lawyer as Soothsayer: Exploring the Important Role of Outcome Prediction in the Practice of Law*, 123 PA. ST. L. REV. 41, 43-44 (2018) (noting that a principal role for lawyers, apart from advocacy, is outcome-prediction as advisor and prognosticator, both because of the fiduciary obligation to act in accordance with a client's interests and for the lawyer's own pecuniary benefit in case selection). The fountainhead of scholarly exploration of outcome prediction is two early articles by Stuart Nagel: Stuart S. Nagel, *Applying Correlation Analysis to Case Prediction*, 42 TEX. L. REV. 1006 (1964); Stuart Nagel, *Using Simple Calculations to Predict Judicial Decisions*, AM. BEHAV. SCIENTIST (1960).

   25 A common way of putting this is that machine learning models learn from "examples rather than instructions." *Machine Learning*, IBM (May 2019), https://www.ibm.com/design/ai/basics/ml [https://perma.cc/BGB4-Q7VD].

   26 A full accounting of machine learning is beyond the scope of this Article. For a lawyer-accessible overview of key concepts (train-test splits, k-fold validation, optimizing bias and variance, overfitting), see Ryan Copus, Ryan Hübert & Hannah Laqueur, *Big Data, Machine Learning, and the Credibility Revolution in Empirical Legal Studies*, *in* LAW AS DATA, *supra* note 8, at 25. A leading textbook treatment is GARETH JAMES, DANIELA WITTEN, TREVOR HASTIE & ROBERT TIBSHIRANI, AN INTRODUCTION TO STATISTICAL LEARNING 183 (2017). For an accessible exploration of causal inference in statistics, see JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE, MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST'S COMPANION (2008).

tech applications use "unsupervised" methods that find patterns in data without pre-labeled examples, leaving to humans to determine post hoc which ones matter.[27] Second, many current legal tech tools leverage conventional ML techniques built around highly flexible statistical models, or combinations of models. These approaches are powerful because they dispense with the rigid, across-the-board assumptions about the functional form of data that characterize and limit conventional data science methods, but they are recognizable to those with quantitative training.[28] Going forward, however, the most advanced legal tech tools are likely to use "neural networks"—inspired by the structure of neurons in the human brain and the most common exemplar of an advanced form of ML referred to as "deep learning"—to perform extremely subtle, multi-layered analyses.[29] In Section I.B. below, we offer a quasi-technical accounting of the possibilities and limits of deep learning applied to natural language processing (NLP), the family of techniques that performs text analytics and so holds the most promise for a discipline like law that trades in words.[30] For now, it is enough to note that ML in all its forms can potentially generate highly accurate predictions where conventional data science may not, and so it is—and is likely to continue to be—the technical guts of the more consequential legal tech tools.[31]

Third, looking across Table 1's entrants reveals a set of technical and operational distinctions that will condition legal tech's trajectory and implications. For instance, legal tech tools vary in the degree to which they draw upon technical versus legal expertise and, relatedly, the stage at which that expertise plugs into the tool's development and use. A case-level outcome prediction tool that aids a litigant's forum-shopping calculus by analyzing a sea of past cases to estimate her relative chances across available jurisdictions may largely pose problems of data science. As we detail in Section II.B, key challenges will be empirical measurement (e.g., how to quantify judge ideology), extracting case features from docket sheets or other texts, and obtaining sufficiently large datasets. Moreover, the technical expertise needed

---

[27] JAMES ET AL., *supra* note 26, at 26, 373; *see also* ASHLEY, *supra* note 14, at 246.

[28] An example of a flexible ML model is a decision tree model. *See* JAMES ET AL., *supra* note 26, at 303; Copus et al., *supra* note 26. Less flexible methods, which might be faster or easier to understand, enlist a computer to search across a predetermined set of ways to make predictions. *Id*.

[29] For an accessible introduction to neural networks, see Victor Zhou, *Machine Learning for Beginners: An Introduction to Neural Networks*, TOWARDS DATA SCI. (Mar. 5, 2019), https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9 [https://perma.cc/E4EQ-BGWN].

[30] *See* Mireille Hildebrandt, *Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics*, 68 U. TORONTO L.J. 12, 27 (Supp. 1, 2018) (noting NLP's centrality to legal tech); Frank Fagan, *Natural Language Processing for Lawyers and Judges*, 119 MICH. L. REV. (forthcoming 2021) (manuscript at 9) (same).

[31] On ML's predictive superiority, see Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Ziad Obermeyer, *Prediction Policy Problems*, 105 AM. ECON. REV.: PAPERS & PROCEEDINGS 491, 492-93 (2015).

to create such a tool may largely feed into an up-front process of software development. Once a software platform has been built, a lawyer need only input key case features—or, as technology advances, perhaps just feed in a complaint and pleadings—to prime it.

Other legal tech tools, in contrast, will require significant lawyerly engagement throughout the design and implementation process. For example, a legal analytics tool that tells lawyers which arguments to advance or avoid in a case before a specific judge will likely require, at least given current technology, substantial lawyer input to construct logical models of doctrinal tests or legal factors that past courts have applied in order to guide, and then iteratively revise, the machine's identification and analysis of relevant case law.[32] Another example, and the starkest contrast from the outcome-prediction tool just described, is the suite of technology-assisted review (TAR) and predictive coding tools increasingly used in discovery in large and complex cases.[33] As discussed in more detail in Section II.A, TAR tools follow a common protocol in which lawyers first perform manual review of a subset of documents—sometimes called a "training set" or "seed set"— to provide the "labeled" data upon which supervised machine learning tools rely.[34] Thereafter, as the machine surfaces documents, lawyers are re-deployed to review documents flagged by the machine and add them to the training set as the system iterates toward a best model.[35]

To be sure, the expertise required to implement a given tool need not be exclusively technical or legal. At least for the moment, TAR tools depend on lawyers, but they also require significant technical expertise, both up front and during implementation. Contrary to popular belief, machine learning models are not merely turned loose on data; rather, programmers make a myriad of decisions about how to partition data, which model types and data features to choose, and how much to tune the model.[36] For now, the take-

---

32 *See* Kevin D. Ashley & Stefanie Bruninghaus, *A Predictive Role for Intermediate Legal Concepts*, *in* LEGAL KNOWLEDGE AND INFORMATION SYSTEMS: JURIX 2003 153 (Daniele Bourcier ed., 2003). For more discussion, see *infra* notes 163–166 and accompanying text.

33 "TAR" is the more general term; "predictive coding" is a marketing term used by only some vendors. In what follows, we tend to use "TAR" as the more inclusive term.

34 For an accessible overview of TAR, see *The Sedona Conference Glossary: E-Discovery and Digital Information Management (4th ed.)*, 15 SEDONA CONF. J. 305, 357 (2014); Seth Katsuya Endo, *Technological Opacity & Procedural Injustice*, 59 B.C. L. REV. 821, 822-24 (2018); Dana A. Remus, *The Uncertain Promise of Predictive Coding*, 99 IOWA L. REV. 1691, 1701-06 (2014); Gideon Christian, *Predictive Coding: Adopting and Adapting Artificial Intelligence in Civil Litigation*, 97 CANADIAN BAR REV. 486, 492-93 (2019).

35 *See supra* note 34. For a more technical accounting, see *infra* notes 116–118 and accompanying text.

36 *See* Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 28 (2017) ("[W]hile the control algorithm is not developed by a human, the learning algorithm, the data, and any necessary guiding hints are."); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L.

home point—returned to in more detail below—is that a tool's ratio of technical to lawyerly expertise and the point at which that expertise plugs into its design and use may shape a tool's effect on the role and status of lawyers within the system, the ability of generalist judges to oversee its use, and its distributive impact as between litigation's haves and have nots.

A final notable operational distinction is that legal tech tools vary in their data inputs and, in particular, whether those inputs are widely available at little or no cost, or instead are proprietary and thus held only by certain actors within the system. Of course, much of the legal system operates in full view, and one might think it provides a treasure trove of constantly updating and curated data as an army of litigants and judges move product through it. However, data limitations will significantly shape legal tech's future. A core challenge for outcome prediction and legal analytics tools is the pervasiveness of "secret settlements" and the fact that most settled cases exit docket sheets via unelaborated voluntary dismissals under Rule 41.[37] Importantly, however, these constraints may affect some system actors more than others. Past representations give large law firms a ready-made source of data—including case outcomes, but also document productions and repositories of contracts— to develop and optimize legal tech tools, subject only to client consent to use them.[38] Other actors within the system who trade in large case volumes— among them insurance companies and litigation financiers—may likewise have privileged access to data and be uninclined to share it.

Mapping the full landscape in this way suggests an entirely different set of frameworks for thinking about legal tech than the current debate's overriding focus on the future health of the legal profession. In so doing, it helps to tee up more expansive thinking about legal tech's trajectory and implications—the subject of Sections I.B and I.C—and ultimately informs Part II's case studies of how civil procedure might adapt in response.

B. Technical Limits and the Trajectory Puzzle

A second key task in taking legal tech's measure is to realistically and concretely forecast its future trajectory. Just how far will Table 1's tools advance

---

REV. 653, 683-700 (2017) (discussing the human interaction involved in statistics review, data partitioning, model selection, and model training).

   37 *See* Osbeck, *supra* note 24, at 99 (noting unavailability of settlement information except in specialty areas—e.g., securities litigation); Nora Freeman Engstrom, Legal Ethics: The Plaintiffs' Lawyer (2019) (unpublished manuscript) (collecting sources estimating secret settlements). Still another barrier is the incompleteness of many commercial legal databases. *See generally* Merritt McAlister, *Missing Decisions*, 169 U. PA. L. REV. 1101 (2021).

   38 *Cf.* Flanagan & Dewey, *supra* note 19, at 1261 (noting potential usability of "internal law-firm information" but also detailing barriers to doing so, including "client confidences"); *see also* Re & Solow-Niederman, *supra* note 3, at 259-60 (noting the problem of "proprietary data sets").

in the near- to medium-term? Though it runs contrary to the futurist orientation of much of the existing literature, the best way to accomplish this is not by imagining robotic endpoints but rather by gauging legal tech's current capabilities and then soberly evaluating the barriers to further advances.

A growing literature starts down that road by exploring the impediments that will condition legal tech's future. A significant regulatory constraint is Model Rule of Professional Conduct 5.4 and state counterparts outlawing unauthorized practice of law (UPL). Because "practice of law" is capaciously defined, UPL rules have the potential to stunt legal tech tools that operate outside the litigation system, such as lawyer-client matching, automated legal advice, and DIY dispute resolution.[39] A vivid example is the trench warfare between state bars and lawyer-client matching system Avvo.[40] Invoking these struggles, some commentators bet on lawyers' guild-like capacity to fend off even the most potent tech innovations,[41] while others see technology as an unstoppable force even for a strong professional monopoly.[42] Still others focus

---

[39] Some suggest that UPL rules may apply to predictive coding. *See* Remus, *supra* note 34, at 1708 ("[A] predictive-coding approach to discovery . . . raises new questions regarding what constitutes the unauthorized practice of law—questions that are not readily answerable in the current framework of unauthorized practice rules.").

[40] *See* Benjamin H. Barton & Deborah L. Rhode, *Access to Justice and Routine Legal Services: New Technologies Meet Bar Regulators*, 70 HASTINGS L.J. 955 (2019) (providing a detailed account of Avvo's legal challenges).

[41] *See, e.g.*, Gillian K. Hadfield, *Legal Barriers to Innovation: The Growing Economic Cost of Professional Control over Corporate Legal Markets*, 60 STAN. L. REV. 1689, 1724-25 (2008) (explaining how regulation precludes an efficient market for innovative legal tools and thus dampens development incentives); Larry E Ribstein, *The Death of Big Law*, 2010 WIS. L. REV. 749, 807-08 ("Licensing laws also constrain the development of legal information products . . . . Without this regulation, firms would have incentives to invest in, for example, software and data that could automate contract drafting or aspects of litigation . . . .").

[42] *See* McGinnis & Pearce, *supra* note 1, at 3057-64 (describing how "the legal services market has largely become de facto deregulated" and "even increased unauthorized practice enforcement" would not prevent the delivery to U.S. consumers of legal services through machine intelligence); *see also* Ray Worthy Campbell, *Rethinking Regulation and Innovation in the U.S. Legal Services Market*, 9 N.Y.U. J.L. & BUS. 1, 1 (2012) (explaining why innovation occurs despite regulation); BARTON, *supra* note 1, at 3 (2015) (noting that many new legal technologies "have grown so large and prevalent that the time to quietly nip them in the bud has passed"). Part of this debate centered on whether increased adoption of legal tech reflects economic pressure from the 2009 downturn or a broader and deeper trend. *See* William D. Henderson, *From Big Law to Lean Law*, 38 INT'L REV. L. & ECON. 5, 14 (2014) ("To survive and thrive in the years to come, firms will increasingly follow Lean Law principles—better, faster, cheaper through collaboration, process engineering and technology— rather than the Big Law model."); Linna, *supra* note 15, at 393 ("The 2008 U.S. recession accelerated changes in the demand for legal services."); SUSSKIND, *supra* note 1 (discussing drivers of legal change); Yoon, *supra* note 22, at 462 (describing litigants response to the recession as drawing "a harder line on their legal expenses"); *see also* Alarie et al., *supra* note 14, at 110-11 (explaining how post-recession client demands "provided the necessary impetus" for law firms to embrace technological advances); Eli Wald, *Foreword: The Great Recession and the Legal Profession*, 78 FORDHAM L. REV. 2051, 2061-62 (2010) ("Increasingly competitive practice conditions in the

on broader professional and cultural barriers, emphasizing the inherent conservatism of lawyers as a profession,[43] their aversion to "mathiness,"[44] or the disconnect between the heavy, up-front, fixed costs necessary to develop many legal tech tools and a legal services industry that remains economically organized around the billable hour and pass-through of case-specific costs to clients.[45]

These barriers are real and substantial. But the most significant determinant of legal tech's trajectory is likely to be technical, and it extends from an inescapable fact: Law "has language at its heart."[46] As a result, many legal tech tools depend on text analytics and, more specifically, a family of ML techniques noted previously called natural language processing (NLP).[47] At a high level of abstraction, NLP aims to identify patterns in human language in ways that facilitate problem-solving. But, as with machine learning more generally, NLP has many tributaries.

The earliest NLP techniques were simple expert systems—i.e., hand-written rules using, for instance, regular expressions to parse text.[48] A second generation relied upon statistical analysis keyed to the frequencies of words appearing in a corpus of documents in order to draw inferences about their content.[49] The current research frontier, and a rapidly advancing one, is a mix

---

market for corporate legal services, accentuated by the economic downturn, are transforming . . . the practice realities, the organization, and the structure of large law firms . . . .").

43   *See* Flanagan & Dewey, *supra* note 19, at 1256 (noting lawyers' "professionally honed risk aversions").

44   *See* Remus & Levy, *supra* note 1, at 540 (citing the "well-documented distaste that many lawyers have for technology and 'mathiness' of any kind").

45   *See* Flanagan & Dewey, *supra* note 19, at 1260-61 ("If clients will not absorb the cost [of new technologies] or if the vendors' pricing model does not easily permit pass-through billing, then cost becomes a more considerable barrier of adoption.").

46   *See* Robert Dale, *Law and Word Order: NLP in Legal Tech*, MEDIUM (Dec. 15, 2018), https://towardsdatascience.com/law-and-word-order-nlp-in-legal-tech-bd14257ebd06 [https://perma.cc/T5WY-XZNU]; *see also* ALFRED DENNING, THE DISCIPLINE OF LAW (1979) ("Words are the lawyer's tools of trade.").

47   *See supra* notes 30–31 and accompanying text; Hildebrandt, *supra* note 30, at 27 (noting NLP's centrality to legal automation).

48   Canonical overviews of NLP include CHRISTOPHER D. MANNING & HINRICH SCHÜTZE, FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING (1999) and DANIEL JURAFSKY & JAMES H. MARTIN, SPEECH AND LANGUAGE PROCESSING (2d ed. 2008).

49   The basic assumption is that each document was generated from a mix of topics, and each topic was generated from a mix of words. Through statistical analysis of word frequencies, an analyst can infer the topic(s) of new documents and deploy those inferences. *See* Joakim Nivre, *On Statistical Methods in Natural Language Processing*, 13 PROC. NORDIC CONF. COMPUTATIONAL LINGUISTICS (2001) (discussing the use of inferences from corpus data to process natural language); David M. Blei, *Probabilistic Topic Models*, COMM. ACM, Apr. 2012, https://cacm.acm.org/magazines/2012/4/147361-probabilistic-topic-models/fulltext [https://perma.cc/498U-VUPE] (detailing the evolution of topic modeling). Note that the frequencies used, as contained in a "term-document matrix," are not just simple counts. Many statistical NLP applications depend on TF-IDF—short for term-frequency/inverse-document frequency—values in which a term's frequency in a document is discounted by its frequency in the full corpus to avoid merely classifying based on the most common words. ASHLEY, *supra* note 14, at 218. Statistical NLP can be either supervised or unsupervised. A

of linguistics and "deep learning" (i.e., neural network) techniques.[50] In a nutshell, deep-learning NLP machines make language computationally tractable by converting words, sentences, documents, or, in the legal context, entire cases into unique vectors, called "embeddings." Each vector can be envisioned as an arrow from the origin to a point that represents the item of interest in a large, n-dimensional space, its magnitude a function of the presence of words, case citations, indexing concepts, or other features.[51] Once this vast vector space has been constructed and human-annotated labels affixed to training materials (again, words, sentences, documents, cases), a sophisticated machine learning model can manipulate the vectors mathematically using large numbers (on the order of billions) of calculations to model relationships between them.[52] With sufficient data and computing power, the system's outputs enable a range of legal tasks, such as identifying relevant or privileged documents, past legal decisions that may be controlling, or, though we will see it is far trickier, the winning argument in a case.

Many of the more specific technical challenges that will shape legal tech's trajectory are generic NLP challenges. As a data science method, machine learning developed alongside increases in computing power and "big data"—defined as larger *quantities* of data, but also *higher-dimension* data (i.e., data with more predictors)—which presented rich analytic possibilities while exposing the shortcomings of conventional econometrics.[53] But textual data

good example of the latter is Latent Dirichlet allocation, or LDA, a topic modelling tool used to cluster words and documents into topics without up-front guidance. *See generally* CHRIS TUFTS, THE LITTLE BOOK OF LDA: AN OVERVIEW OF LATENT DIRICHLET ALLOCATION AND GIBBS SAMPLING, https://ldabook.com/index.html [https://perma.cc/DG68-8GTS] (describing LDA); David M. Blei, Andrew Y. Ng & Michael I. Jordan, *Latent Dirichlet Allocation*, 3 J. MACHINE LEARNING RES. 993 (2003) (same).

[50] *See* Christopher D. Manning, *Computational Linguistics and Deep Learning*, 41 COMPUTATIONAL LINGUISTICS 701-02 (2015) (describing the "tsunami"-like impact that deep learning will have on NLP). A textbook-length treatment is Jurafsky & Martin, *supra* note 48.

[51] *See* Ashley, *supra* note 22, at 1121 (explaining the vector concept). For technical and applied examples, see Benjamin R. Baer, Skyler Seto & Martin T. Wells, *Exponential Family Word Embeddings: An Iterative Approach for Learning Word Vectors*, 32d CONF. NEURAL INFO. PROCESSING SYS., MONTRÉAL, CAN. (2018). *See also* Elliott Ash, Daniel L. Chen & Arianna Ornaghi, *Gender Attitudes in the Judiciary: Evidence from U.S. Circuit Courts* (Warwick Econ. Rsch. Papers, Paper No. 1256, 2020), https://warwick.ac.uk/fac/soc/economics/research/workingpapers/2020/twerp_1256_-_ornaghi.pdf [https://perma.cc/9BZM-H9VK] (using word embedding to study how gender stereotypes affect the behavior of U.S. Appellate Court judges).

[52] For a good but technical overview, see Tomas Mikolov, Kai Chen, Greg Corrado & Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space (Sept. 7, 2013) (unpublished manuscript), http://arxiv.org/abs/1301.3781 [https://perma.cc/6CZF-WYNK]. Vector space similarity uses a measure of Euclidian distances between the end-points of the vectors in the n-dimensional vector space. By computing the cosine of the angle between a pair of vectors, one can quantify the similarity of vector pairs. The smaller the cosine/angle, the greater the similarity. *Id.* at 5.

[53] *See* Marion Dumas & Jens Frankenreiter, *Text as Observational Data*, *in* LAW AS DATA, *supra* note 8, at 61 (noting how computational tools allow for representation of texts as "high-dimensional vectors").

brings further unique challenges. The easiest to see arise from the richness of human language.[54] Sarcasm, implicit meanings, multiple words with the same meaning (synonymy), and the same word with multiple meanings (polysemy) are just the beginning.[55] The result is that advanced NLP requires extensive manipulation of raw texts before analytics can be performed. NLP machines must first break text down into manipulable pieces by normalizing and tokenizing it (i.e., eliminating superficial variations in words via "stemming," and removing punctuation and "stop words"[56]), parsing it (i.e., tagging words for parts of speech and other syntactic structure, including grammatical roles), and representing it (i.e., converting the reduced form "tokens" to vector-based embeddings that permit semantic comparisons).[57] This latter step relies upon an encoder-decoder that assigns semantic value to a word based on its context—that is, the words appearing before and after it—to disambiguate it and link it to synonyms to move its representation closer to its intended human meaning.[58]

---

[54] Adam Zachary Wyner, *Weaving the Legal Semantic Web with NLP*, CORNELL LEGAL INFO. INST. (May 17, 2010, 2:27 PM), https://blog.law.cornell.edu/voxpop/2010/05/17/weaving-the-legal-semantic-web-with-natural-language-processing [https://perma.cc/N84L-84V5] (discussing the intricacies of natural language).

[55] A more technical framing is that language's large lexicon, rich grammar, and near-infinite semantic realizations renders text analytics a sparse and underdetermined problem.

[56] Stop words are high-frequency words with "low information content," such as articles or pronouns. *See* Peter D. Turney & Patrick Pantel, *From Frequency to Meaning: Vector Space Models of Semantics*, 37 J. A.I. RSCH. 141, 154 (2010) (explaining the methods by which NLP machines process pronunciation, grammar and syntax).

[57] For an overview of the first two steps, see Elvis Saravia, *Fundamentals of NLP—Chapter 1—Tokenization, Lemmatization, Stemming, and Sentence Segmentation*, NOTEBOOKS BY DAIR.AI, https://dair.ai/notebooks/nlp/2020/03/19/nlp_basics_tokenization_segmentation.html [https://perma.cc/XA46-WWTG].

[58] Attention is where the most rapid recent advances in NLP have come, particularly the 2018 publication of Google's BERT model. *See* Jacob Devlin, Ming-Wei Chang, Kenton Lee & Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (May 24, 2019) (unpublished manuscript), https://arxiv.org/abs/1810.04805 [https://perma.cc/V5S7-2GEF] (discussing the means by which BERT achieves "bi-directional representations from unlabeled text"); Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser & Illia Polosukhin, Attention Is All You Need (Dec. 6, 2017) (unpublished manuscript), https://arxiv.org/abs/1706.03762 [https://perma.cc/E2HV-WCPS] (proposing a new network architecture that uses attention mechanisms to achieve language representations of superior quality). An accessible, non-math explanation of vectorization as contextualization and a solution to synonymy, etc., is Noah A. Smith, *Contextual Word Representations: Putting Words into Computers*, COMM. OF THE ACM, June 2020, at 66. Another significant recent advance is Open AI's GPT-3 model. Like BERT, it is a self-supervised learning model that pre-trains a base model by obfuscating a word from a sentence and training the model to predict that word. In a sense, the vast corpus of human language, not humans explicitly engaged in labeling, provides the labels. Once the base model is trained, it can be "finetuned" or "primed" for more specific tasks on smaller datasets. On self-supervision and also GPT-3's specific approach, see Longlong Jing & Yingli Tian, Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey (Feb. 16, 2019) (unpublished manuscript), https://arxiv.org/abs/1902.06162 [https://perma.cc/D5GS-68SE]; Tom B.

A related challenge is that advanced NLP is computationally demanding. The most cutting-edge applications require enormous compute power to perform the billions of calculations required for even seemingly straightforward tasks. In one sense, NLP's cost has declined recently due to the availability of open-source tools (e.g., Google's TensorFlow software[59] and its BERT encoder-decoder system,[60] Stanford's CoreNLP,[61] and Facebook's PyTorch[62]). There is, however, a potential trade-off. Open-source, off-the-shelf NLP models are trained on general text corpora (e.g., Wikipedia, the so-called Google Books "corpus,"[63] IMDb movie reviews), and their language representations may not "transfer" well to domain-specific, technocratic areas, particularly "legalese."[64] For many discrete legal tasks, fully harnessing NLP may thus require significant re-training of pretrained models—and may also require data and computing power that tends to be concentrated in key industrial players, such as law firms and tech companies.[65] While domain adaptability remains an open research question in computer science, the need for retraining to improve upon benchmark NLP tasks could be a significant constraint.

The most acute challenge facing legal tech is more law-specific: NLP cannot yet reliably "read" legal texts in the sense of extracting legal concepts

Brown et al., Language Models Are Few-Shot Learners (July 22, 2020) (unpublished manuscript), https://arxiv.org/abs/2005.14165 [https://perma.cc/8MAT-WCCN].

   59 *See* TENSORFLOW, https://www.tensorflow.org [https://perma.cc/NK2V-G4ZM].

   60 *See* Devlin et al., *supra* note 58.

   61 *See* CORENLP, *https://stanfordnlp.github.io/CoreNLP/* [https://perma.cc/EE3D-G22W]; *see also* Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard & David McClosky, *The Stanford CoreNLP Natural Language Processing Toolkit*, 52 PROC. ANN. MEETING ASS'N COMPUTATIONAL LINGUISTICS: SYSTEM DEMONSTRATIONS 55 (2014).

   62 *See Tools: PyTorch*, FACEBOOK AI, https://ai.facebook.com/tools/pytorch [https://perma.cc/3UJS-R8PR].

   63 Eitan Adam Pechenick, Christopher M. Danforth & Peter Sheridan Dodds, *Characterizing the Google Book Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution*, PLOS ONE (Oct. 7, 2015), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137041 [https://perma.cc/6EUS-PZ6W].

   64 *See* Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolas Aletras & Ion Androutsopoulos, LEGAL-BERT: The Muppets Straight Out of Law School (Oct. 6, 2020) (unpublished manuscript), https://arxiv.org/abs/2010.02559 [https://perma.cc/4Z6Q-23T3] (detailing the comparatively poor performance of the model used in the legal context without the proper model pre-training). For transfer learning challenges in other expert domains, see Iz Beltagy, Kyle Lo & Arman Cohan, SciBERT: A Pretrained Language Model for Scientific Text (Sept. 10, 2019) (unpublished manuscript), https://arxiv.org/abs/1903.10676 [https://perma.cc/57Y5-GPHA] (finding that a domain specific-model over scientific texts outperformed a model trained on general texts); Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann & Matthew McDermott, *Publicly Available Clinical BERT Embeddings*, 2 PROC. CLINICAL NAT. LANGUAGE PROCESSING WORKSHOP 72 (2019), https://www.aclweb.org/anthology/W19-1909 (finding that a domain specific-model over clinical texts outperformed a model trained on general texts).

   65 *See* Flanagan & Dewey, *supra* note 19, at 1259 (noting transfer problem in legal context); Remus & Levy, *supra* note 1, at 522 (providing accessible overview of language parsing tools). For the privileged access of litigation's "haves" to needed data, see *infra* notes 106–109 and accompanying text.

or legal rules in logical forms.[66] One reason is that, while second-nature to seasoned lawyers, legal reasoning consists of a dizzying array of analytic moves. Case outcomes often turn on a dense mix of rule-based reasoning and case-based reasoning, including: linguistic arguments about a statutory or regulatory term's "ordinary" meaning; systemic arguments about harmonization across statutory sections; analogical arguments from past case law; evidentiary arguments about key facts; and teleological arguments from legislative purposes or other substantive values.[67] Not only must the machine identify and manipulate different types of legal argument—linguistic, systemic, analogical, evidentiary, teleological—it must also develop traffic rules for navigating between them.[68]

To be sure, NLP has improved rapidly in its capacity to parse legal argument. NLP can now identify the rhetorical roles played by sentences in court decisions (e.g., statements of legal rules, fact determinations) and who among possible speakers (judge, litigants, testifying expert, evidentiary document) is making an assertion.[69] Classification and attribution of this sort

---

[66] *See* Ashley, *supra* note 22, at 1120, 1136("Computer programs cannot yet read legal texts like lawyers can. . . . In particular, computer programs cannot read contracts the way that attorneys do."); *see also* ASHLEY, *supra* note 14, at 3 (distinguishing between "legal information retrieval" and "argument retrieval").

[67] ASHLEY, *supra* note 14, at 88.

[68] *Id*. Law is also indeterminate because legal language is ambiguous semantically (e.g., "reasonable" and "discrimination," which both suffer from uncertainty about the boundaries of what the terms refers to) and syntactically (the logical connectors—the "ands" and "ors"—that structure propositions). Edward Levi, *An Introduction to Legal Reasoning*, 12 U. CHI. L. REV. 501 (1948). Computational modeling requires "propositionalizing" legal rules in ways that reduce both forms of ambiguity and then weight each proposition.

[69] *See, e.g.*, Vanessa Wei Feng & Graeme Hirst, *Classifying Arguments by Scheme*, 49 PROC. ANN. MEETING ASS'N FOR COMPUTATIONAL LINGUISTICS 987 (2011) (demonstrating an experiment's relative success in identifying common argumentation structures); Ben Hachey & Claire Grover, *Extractive Summarization of Legal Texts*, 14 A.I. & L. 35 (2006) (detailing research project on the automatic text summarization of rulings from the UK House of Lords); Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau & Chris Reed, *Automatic Detection of Arguments in Legal Texts*, 11 PROC. INT'L CONF. A.I. & L. 225 (2007) (detailing results of experiments on detection of arguments in legal texts); M. Saravanan, B. Ravindran & S. Raman, *Improving Legal Information Retrieval Using an Ontological Framework*, 17 A.I. & L. 101 (2009) (explaining the benefits of ontological frameworks to extract legal judgments); Vern R. Walker, Parisa Bagheri & Andrew J. Lauria, Argumentation Mining from Judicial Decisions: The Attribution Problem and the Need for Legal Discourse Models (2015) (unpublished manuscript peer-reviewed and presented at the 2015 Workshop on Automated Detection, Extraction & Analysis of Semantic Info. in Legal Texts), https://sites.hofstra.edu/vern-walker/wp-content/uploads/sites/69/2019/12/WalkerEtAl-AttributionAndLegalDiscourseModels-ASAIL2015.pdf [https://perma.cc/5E4P-5ALJ] (detailing the outstanding challenges of identifying augmentation structures using automated methods). One reason this is hard is that legal texts, relative to other kinds of texts, lack a common structure and format. For a proposal that judges write "structured, machine-readable" opinions and tag "key elements . . . treating facts, issues, cited cases or other elements of the case as pieces of data," see Jameson Dempsey & Gabriel Teninbaum, *May It Please the Bot?*, MIT COMPUTATIONAL L. REP. (Aug. 14, 2020), https://law.mit.edu/pub/mayitpleasethebot/release/1 [https://perma.cc/2DM8-G2AC].

are canonical NLP tasks and a critical step for analyzing legal texts.[70] But NLP has not yet made the leap from these simpler tasks to full-on argument mining—that is, automated discovery of discourse structure and argument-related information, including propositions, premises, conclusions, and exceptions.[71] This is important, because argument representations serve as the bridge between legal texts and a wide range of legal cognitions to which legal tech aspires, from information retrieval and legal analytics to outcome prediction.[72] For each of these tasks, it is only with a jump to fully computational analysis of substantive legal merits that legal tech can perform tasks with robust reasoning and thus *explain* machine outputs in ways that a lawyer can put to use or a client or judge might expect.[73] A machine prediction that a case has an 80 percent chance of victory might help a lawyer decide whether to file a complaint or seek an early settlement; however, it tells her precisely nothing actionable about how to actually win the case.[74]

A consequence is that legal tech tools are currently bounded by their supervised nature—that is, by their need for labeled, typically *lawyer*-labeled, data.[75] For the moment, even the most cutting-edge legal analytics tools require lawyers to perform two critical and resource-intensive tasks. First, lawyers must *translate* an operative doctrinal test into a hierarchical structure of pre-defined elements—for instance, a list of factors that appear in past cases adjudicating, say, the line between employees and independent

---

[70] On text summarization techniques, see Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez & Krys Kochut, Text Summarization Techniques: A Brief Survey (July 28, 2017) (unpublished manuscript), https://arxiv.org/pdf/1707.02268.pdf [https://perma.cc/B4Y5-KXY2].

[71] *See* Ashley, *supra* note 22, at 1117 (implying that NLP has yet to achieve legal information retrieval to "match document structure, concepts, and argument roles with aspects of the problems users seek to solve"); *see also* Remus & Levy, *supra* note 1, at 538 (describing still-embryonic field of argument mining).

[72] *See* Faraz Dadgosari, Mauricio Guim, Peter A. Beling, Michael A. Livermore & Daniel N. Rockmore, *Modeling Law Search as Prediction*, A.I. & L. (2020) (discussing the importance of arguments in information retrieval to complete legal tasks).

[73] *See* ASHLEY, *supra* note 14, at 18; Ashley, *supra* note 22, at 1138. *See generally* Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633 (1995) (discussing explanation in law).

[74] ASHLEY, *supra* note 14, at 23 (noting that fully functional computational legal analysis requires "an ability to explain its reasoning, and that reasoning has to be intelligible to legal practitioners"); Remus & Levy, *supra* note 1, at 550 (noting that most prediction programs "give a user results without showing the precise combination of factors that produced those results").

[75] *See* ASHLEY, *supra* note 14, at 5 (noting that computer science has not yet learned to "automate the knowledge representation process"). *See also* Talley, *supra* note 1, at 185 (noting law's "irreducible complexity" and concluding that it "will necessarily implicate significant human input over the longer term"). For an interesting account of big tech's investment in supervision and the rise of the "labeling" industry, see Madhumita Murgia, *AI's New Workforce: The Data-Labelling Industry Spreads Globally*, FIN. TIMES (July 23, 2019), https://www.ft.com/content/56dde36c-aa40-11e9-984c-fac8325aaa04 [https://perma.cc/7DKJ-UJQZ].

contractors.[76] Second, lawyers must *annotate* legal texts in order to train machines to identify these argument-related elements—here again, legal factors or other discourse structures—in old cases in order to compare them to new ones.[77] The results of this lawyer-intensive process of translation and annotation can be powerful. Fed well-labeled data, machine learning tools can determine that factor X or an entire case, long thought to drive case outcomes, has become, or has always been, irrelevant.[78] Put another way, legal tech tools perform well in assigning weights to legal factors, even if they cannot, as of yet, discover those factors on their own.[79] But the result is still a long way from the fully automated robolawyers and robojudges in the more futurist accounts of legal tech.[80] It is only with significant further NLP advances that legal tech will achieve the holy grail: a legal app that can, in fully automated fashion, construct an "ontology" of a legal area, extract substantive legal

---

[76] ASHLEY, *supra* note 14, at 11, 209 (noting legal tech's "knowledge acquisition bottleneck" and need for "hand-tooled knowledge representations"); Osbeck, *supra* note 24, at 54 (describing the "element-focused analysis" and "factor tests" that commonly structure much legal reasoning).

[77] Even then, detection of argumentative structures is far from perfect. *See, e.g.*, Kevin D. Ashley & Stefanie Bruninghaus, *Automatically Classifying Case Texts and Predicting Outcomes*, 17 A.I. & L. 125 (2009) (describing results of the SMILE interface for extracting argument-related legal factors in legal texts but noting reliability concerns); Adam Wyner & Wim Peters, *Semantic Annotations for Legal Text Processing Using GATE Teamware*, 2012 PROC. LANGUAGE RES. & EVAL. CONF.: SEMANTIC PROCESSING OF LEGAL TEXTS WORKSHOP 34 (describing an annotation pipeline for identifying legal factors in court decisions); Adam Wyner & Wim Peters, *Towards Annotating and Extracting Textual Legal Case Factors*, 2010 PROC. LANGUAGE RES. & EVAL. CONF.: SEMANTIC PROCESSING OF LEGAL TEXTS WORKSHOP 36 (same).

[78] *See* Fagan, *supra* note 30, at 13 ("Old-fashioned positive studies of black-letter law carried out with sophisticated NLP tools can clarify standards and help make law more precise."); Wolfgang Alschner, *AI and Legal Analytics*, *in* AI AND THE LAW IN CANADA (Teresa Scassa & Florian Martin-Bariteau eds., forthcoming 2021) (manuscript at 4) (noting the potential of network analysis models to "empirically support or debunk textbook accounts of what judicial decisions proved most influential"); Jens Frankenreiter & Michael Livermore, *Computational Methods in Legal Analysis*, 16 ANN. REV. L. & SOC. SCI. 39 (2020) (reviewing NLP uses for "interpretation and description" of doctrinal trends). For recent concrete examples, see Jonathan Macey & Joshua Mitts, *Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil*, 100 CORNELL L. REV. 99, 115 (2014); Frank Fagan, *From Policy Confusion to Doctrinal Clarity: Successor Liability from the Perspective of Big Data*, 9 VA. L. & BUS. REV. 391, 394-95 (2015); Frank Fagan, *Waiving Good Faith: A Natural Language Processing Approach*, 16 NYU J.L. & BUS. 633, 633 (2020); Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro & Vasileios Lampos, *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, PEERJ COMPUT. SCI. (Oct. 24, 2016), https://peerj.com/articles/cs-93.pdf [https://perma.cc/AQ28-JYER]. For an early effort to computationally model "conceptual change" in case law, see Edwina L. Rissland & Timur M. Friedman, *Detecting Change in Legal Concepts*, 5 PROC. INT'L CONF. ON A.I. & L. 127 (1995).

[79] *See* ASHLEY, *supra* note 14, at 125 (noting distinction between assigning weights to factors and discovering them).

[80] *See supra* notes 7–13 and accompanying text (outlining the possible future in which the law becomes self-driving); *see also* Remus & Levy, *supra* note 1, at 521 (noting that the closest legal tech has come to retrieving underlying arguments are the Q/A—i.e., question-and-answer—systems touted by Ross Intelligence and others that purport to retrieve case passages in response to natural language questions).

features from relevant legal texts, and then link those features to computational models to perform key tasks, from retrieving well-tailored legal materials to predicting case outcomes.[81]

Finally, even with significant advances in NLP, it is possible that legal tech tools will not, or at least not soon, be able to mimic the legal cognitions that seasoned lawyers possess. There is no single way to capture what lawyers do, but a rough approximation is that sound legal judgment must be both synoptic and subtle.[82] It must be synoptic in simultaneously marrying an "internal" perspective on a case (grappling with law on its own terms and under its own logic—i.e., its legal "merit") with an "external" perspective (how a particular judge or litigant or myriad other case characteristics external to a case's internal logic or "merit" relate to the outcome).[83] Legal judgment must also be subtle in its capacity to parse highly individualized, near-infinite fact patterns, work back and forth between minor fact shadings and legal propositions, sift holdings and dicta, transport concepts from one legal area to another, and account for policy-based and equitable "teleological" reasons.[84] Legal judgment further depends on an ability to predict subtle changes over time, particularly social (and thus judge- or jury-held) norms.[85] Indeed, law's dynamism means that weighing both the "internal" and

---

[81] ASHLEY, *supra* note 14, at 172, 354.

[82] Another helpful framing reduces legal cognition to two problems: haystack problems (i.e., assembly of relevant "needles" from a haystack of materials) and forest problems (extracting overall trends and themes and/or weighing the "gravitas" of particular trees—a form of dimensionality reduction). *See* Vlad Eidelman, Brian Grom & Michael A. Livermore, *Analyzing Public Comments*, *in* LAW AS DATA, *supra* note 8, at 235, 257.

[83] *See* Michael A. Livermore & Daniel N. Rockmore, *Distant Reading the Law*, *in* LAW AS DATA, *supra* note 8, at 8-9. *See also* ASHLEY, *supra* note 14, at 107 (offering a similar internal-versus-external accounting); Katz, *supra* note 1, at 962 (same). For an important early effort that offers a similar "internal" and "external" account of prediction using the contrasting views of Justice Holmes and Karl Llewellyn, see Frederick Schauer, *Prediction and Particularity*, 78 B.U. L. REV. 773 (1998).

[84] *See* Remus & Levy, *supra* note 1, at 549 (describing how, while computers can serve predictive functions, they cannot yet move the law forward through creative ways). For more on case equities, see Frank Pasquale & Glyn Cashwell, *Four Futures of Legal Automation*, 63 UCLA L. REV. DISCOURSE 26, 45 (2015); Trevor Bench-Capon & Giovanni Sartor, *A Model of Legal Reasoning with Cases Incorporating Theories and Values*, 150 A.I. 97 (2003); Donald H. Berman & Carole D. Hafner, *Representing Teleological Structure in Case-Based Legal Reasoning: The Missing Link*, 4 PROC. INT'L CONF. ON A.I. & L. 50, 50-59 (1993); Alison Chorley & Trevor Bench-Capon, *AGATHA: Automated Construction of Case Law Theories Through Heuristic Search*, 10 PROC. INT'L CONF. ON A.I. & L. 45-54 (2005).

[85] *See* Levi, *supra* note 68, at 501-04 (describing law as a "moving classification scheme"). This is especially true with case-based reasoning, where courts working against a precedential backdrop must decide whether to restrict, extend, or replace legal concepts to deal with new and proximate fact contexts or shifting social values—a process that often turns on teleological considerations. *See* ASHLEY, *supra* note 14, at 80 ("The challenge for cognitive computing is how to design computer programs that can assist users in constructing such arguments by formulating theories, linking them to analogous positive case examples and distinguishing them from negative instances."); *see also* KARL BRANTING, REASONING WITH RULES AND PRECEDENTS 9-25 (1999) (describing approach that isolates "criterial facts" driving judicial decisions).

"external" determinants of a case is more than just a brute-force analytic exercise. It is a subtle, trend-sensitive, predictive one.

When it comes to synopticism, computation may well prove superior to human cognition. Indeed, computation's comparative advantage may be its ability to perform sweeping and comprehensive analyses of myriad case factors, whether "internal" or "external" to law, and perfectly weight each.[86] An insightful way to capture legal tech's promise in this regard, highlighted by Livermore and Rockmore, is that NLP permits a kind of "distant reading"—an idea lifted from literary criticism to describe analysis of large text corpora at a coarse level of abstraction, as contrasted with the "close reading" literary critics and lawyers perform.[87]

The more difficult question may be whether machines can perform legal judgment's subtler analytic tasks as well as or better than humans. Here, there is more reason to doubt machine prowess. Machine-based legal analytics, focused as they are on "distant" or coarse pattern recognition, may only be able to handle easy cases, not hard ones,[88] and may miss subtly evolving internal (doctrinal) or external (social) trends.[89] Human cognition, the argument goes, is strongest in its capacity for parsimonious reasoning with incomplete information—precisely the cases at the doctrinal frontier where fine-grained fact distinctions or less tractable "teleological" arguments control.[90]

A pair of conclusions follow from this quasi-technical accounting of legal tech's possibilities and limits. First, legal tech's advance will not be

---

[86] *See* Alschner, *supra* note 78 (manuscript at 5) (comparing the ability of a computational tool to comprehensively consider past caselaw to the "convenience sample" of human analysis).

[87] *See* Livermore & Rockmore, *supra* note 83, at 8-9; *see also* FRANCO MORETTI, GRAPHS, MAPS, TREES: ABSTRACT MODELS FOR A LITERARY HISTORY 1 (2005) (coining the term in the literary criticism context); Mireille Hildebrandt, *The Meaning and Mining of Legal Texts*, *in* UNDERSTANDING DIGITAL HUMANITIES: THE COMPUTATIONAL TURN AND NEW TECHNOLOGY 145, 153 (David. M. Berry ed., 2012).

[88] *See* Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2003 (2019) (noting human adjudication's perhaps unique "facility for 'hard cases'"); *see also* Pasquale & Cashwell, *supra* note 84, at 43 (describing how human judgment is needed where more complexity is present, such as with persuasive authority).

[89] Remus and Levy frame this latter point as ML's limited ability to handle "unanticipated contingencies." Remus & Levy, *supra* note 1, at 538, 541.

[90] Alarie, *supra* note 4, at 444; *see also* ASHLEY, *supra* note 14, at 12 (explaining the paradigms of IR that aid attorneys in the parsing of information necessary to solve a problem rather than replace human thinking); Frank Pasquale & Glyn Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviourism*, 68 U. TORONTO L.J. 63, 68 (2018) (noting the argument that facts matter more than law in case outcomes, rendering predictive AI software in this area weaker than a human's cognitive capability). The counter from computer scientists is that parsimonious reasoning under limited information is just a form of "dimensionality reduction"—separating wheat from chaff—and a task that computers can and will perform better than humans. Even grasping trending social dynamics may ultimately favor machines because messy social processes tend to be enshrined in speech that NLP can access more efficiently. *See* Dumas & Frankenreiter, *supra* note 53, at 67 (outlining how big data provides opportunities to study social processes).

monolithic. Rather, its incorporation into the civil justice system will be siloed, incremental, and halting—across prediction tasks and subject-matter silos.[91] In particular, it is a good bet that legal tech tools will arrive sooner, and advance most rapidly, in legal areas where data is abundant, regulated conduct takes repetitive and stereotypical forms, legal rules are inherently stable, and case volumes are such that a repeat player stands to gain financially by investing.[92] This helps to explain why some of the most advanced legal tech tools are found in technocratic and self-contained areas of law (e.g., tax, labor and employment, patents), or highly routine ones (e.g., auto accidents), but not more open-ended legal contexts.[93] The question is whether, or how quickly, prediction tools can move beyond those self-contained legal areas, and how soon it will reliably perform other higher-order legal cognitions, including legal search and analysis that goes beyond simple "hunting and

---

[91] Remus & Levy, *supra* note 1, at 511-37 (offering predictions about which legal tasks will be subject to "light, moderate, or heavy employment effects").

[92] *See* Casey & Niblett, *supra* note 11, at 5 (predicting the "likely rise of a market for third-party vendors providing certified computer code to govern contractual relationships"); *see also* Remus & Levy, *supra* note 1, at 538 (noting "major inroads" in some areas, especially e-discovery, but stating that legal tech tools remain "embryonic" in other areas, particularly legal analytics); John Armour & Mari Sako, *AI-Enabled Business Models in Legal Services: From Traditional Law Firms to Next-Generation Law Companies?*, 7 J. PROS. & ORG. 27, 30 (2020) (noting that legal successes have "so far been limited to large organizations with sufficient value at stake to justify the investment"). *See generally* ASHLEY, *supra* note 14, at 72 (discussing "standardized schemes . . . developed for annotating or tagging statutes and regulations with procedural and substantive semantic information that can then be used to search for relevant provisions").

[93] *Compare* Alarie, *supra* note 4, at 446 (predicting success with regard to tax law), Benjamin Alarie, Anthony Niblett & Albert H. Yoon, *Using Machine Learning to Predict Outcomes in Tax Law*, 58 CANADIAN BUS. L.J. 231, 238 (2016) (reporting success in the tax area), Benjamin Alarie & Abdi Aidid, *Predicting Economic Substance Cases with Machine Learning*, J. TAX PRAC. & PROC., Summer 2020, at 35, 36 (demonstrating that machine learning "algorithms can correctly predict the case outcomes and give practitioners opportunities to refine their understanding and ultimately provide better tax advice"), Blakeley McShane, Oliver P. Watson, Tom Baker & Sean J. Griffith, *Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits*, 9 J. EMPIRICAL LEGAL STUD. 482 (2012) (same as to securities), Mihai Surdeanu, Tamesh Nallapati, George Gregory, Joshua Walker & Christopher D. Manning, *Risk Analysis for Intellectual Property Litigation*, 13 PROC. INT'L CONF. A.I. & L. 116 (2011) (reporting a predictive accuracy rating of 64% in patent cases), Hannes Westermann, Vern R. Walker, Kevin D. Ashley & Karim Benyekhlef, *Using Factors to Predict and Analyze Landlord-Tenant Decisions to Increase Access to Justice*, 17 PROC. INT'L CONF. ON A.I. & L. (2019) (reporting significant success in landlord-tenant disputes), *and* Samuel Dahan, Jonathan Touboul, Jason Lam & Dan Sfedj, *Predicting Employment Notice Period with Machine Learning: Promises and Limitations*, 65 MCGILL L.J. (forthcoming 2020) (manuscript at 1-2), https://ssrn.com/abstract=3595769 (same as to employment disputes), *with* Charlotte S. Alexander, Khalifeh al Jadda, Mohammad Javad Feizollahi & Anne M. Tucker, *Using Text Analytics to Predict Litigation Outcomes*, *in* LAW AS DATA, *supra* note 8, at 310 (describing prediction challenges in the employment discrimination context). For general discussion, see Ashley, *supra* note 22, at 1123. For a seminal outcome prediction effort, since superseded by powerful new methods, see Theodore W. Ruger, Pauline T. Kim, Andrew D. Martin & Kevin M. Quinn, *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150, 1152 (2004).

gathering" to return cases based on argumentation structure or legal analytics that can identify the best argument to lay before *this* judge.[94]

Second, understanding legal tech's current technical limits suggests that, in the near to medium-term, even the most advanced legal tech tools will entail substantial lawyer engagement. Rather than full automation, legal tech may instead yield a kind of "advanced lawyering"—a spin on chess-master Gary Kasparov's notion of "advanced chess," in which human and machine ally and compete against other human-machine teams, working symbiotically, rather than merely pitting human against machine.[95] Lawyers, on this view, may often use commoditized systems that fully substitute for human lawyering.[96] But a large slice of legal tech will for the foreseeable future remain customized and operate within a paradigm of "cognitive computing" defined by intensive human-machine collaboration, not simple keystrokes, as overseen by a new breed of "hybrid" legal professional.[97]

C. *Implications*

While legal tech's precise technical trajectory is unknowable, the field has nonetheless begun to sketch a set of claims, though often abstract and conflicting ones, about legal tech's likely impacts on lawyers, law, and the legal system. Three concerns predominate: (i) legal tech's effect on the legal profession; (ii) its effect on conceptions and implementations of rule of law; and (iii) its distributive effects. Much of this discussion is jurisprudential, with eyes cast out at a distant horizon populated by robojudges and robolawyers and featuring fully automated decisionmaking. But armed with Section I.A's overview of the legal tech toolkit and Section I.B's sober account of its technical trajectory, we can begin to distill a set of more concrete ways legal tech will impact the litigation system over the near- to medium-term.

---

94 Katz, *supra* note 1, at 957 (stressing the need for the retrieval of analogically similar cases for "highest end prediction").

95 *See, e.g.*, Mary (Missy) Cummings, *Man versus Machine or Man + Machine?*, IEEE INTELLIGENT SYS., Sept./Oct. 2014, at 62, 67 (recognizing that although engineers prefer autonomy to be left to machines, human judgment is necessary for cognitive reasoning machines where "knowledge-based behaviors and expertise are required"); Livermore & Rockmore, *supra* note 8, at xiv; Yoon, *supra* note 22, at 466 ("Intelligence augmentation . . . reflects a symbiotic relationship between humans and technology."); Wu, *supra* note 88, at 2004 ("[F]or the foreseeable future, software systems that aim to replace systems of social ordering will succeed best as human–machine hybrids, mixing scale and efficacy with human adjudication for hard cases.").

96 SUSSKIND, *supra* note 1 (comparing commoditized and customized/bespoke tools).

97 ASHLEY, *supra* note 14, at 35, 350, 355-56 (noting likely ascendance of tools that "engage users in collaboratively posing, testing, and revising hypotheses about how an issue should be decided"); Armour & Sako, *supra* note 92, at 29 (reviewing literature on how different business models will fuel the emergence of "hybrid professionals" and "organizing professionals" in an increasingly automated and digitized legal system).

1. Legal Tech and the Legal Profession

"Predictions of structural change in the legal industry," Michael Simkovic and Frank McIntyre recently noted, "date back at least to the invention of the typewriter."[98] But this has not stopped commentators from weighing in on legal tech's effect on the legal profession. The result is a welter of competing claims running the gamut from continuation of business as usual, with only modest shifts to the traditional set-up of law firms selling billable-hour legal services in a leveraged partner-associate hierarchy, to the near-complete effacement of lawyers by robotic stand-ins. But whatever legal tech's effect on the economic and organizational structure of the legal services industry, most agree that its proliferation will reshape—and, indeed, has already begun to do so—the professional status and authority of lawyers.

Core to the debate over the future of lawyers—but also emblematic of its unsettled nature—is the application of a standard pair of economics concepts: Are legal tech and analog lawyering substitutes or complements? On the one hand, it is hard to deny that legal tech will function, to at least some degree, as a substitute for conventional legal services, thus shrinking the profession and reducing aggregate lawyer income (even if it increases the income of law firm equity holders). Big firm lawyers now spend perhaps less than 5% of their time on document review—previously a profit center for the profession's upper echelon.[99] TAR may shrink this further. On the other hand, legal tech and human lawyering can also act as complements, increasing demand for, and thus the premium on, higher-order lawyer judgment, from parsing machine-distilled "hot docs" to crafting litigation strategy. Though some lawyers will be displaced, law practice for the remainders may be both more stimulating and more profitable.[100] Finally, many analyses of displacement miss the fact that the supply of and demand for legal representation are endogenous to its cost. The cheaper legal representation is, the more of it litigants can afford, opening new and potentially profitable markets for legal services to those with cognizable claims who currently, lacking willing counsel, choose to "lump it."[101] Some lawyers will be displaced, but others will

---

98 Michael Simkovic & Frank McIntyre, *The Economic Value of a Law Degree*, 43 J. LEG. STUD. 249, 275 (2014).

99 *See* Remus & Levy, *supra* note 1, at 508 tbl.1 (finding large-firm lawyers spent 4.1% of billable hours on document review from 2012-2015).

100 Casey & Niblett, *supra* note 10, at 1446-47 (noting that similar previous technological advancements reduced transaction costs like the introduction of AI could); Livermore & Rockmore, *supra* note 8, at xiv; Remus & Levy, *supra* note 1, at 533-37 (outlining the likely employment effects from increased use of computers); *see also* Cummings, *supra* note 95, at 62 (noting the possibility of AI justifying replacing humans with automation).

101 *See* Yoon, *supra* note 22, at 470 ("The client benefits from paying smaller legal fees than she would without the technology."); *see also* Remus & Levy, *supra* note 1, at 535 (noting possibility that

find entirely new markets for their skills. PeopleLaw, the steadily shrinking sector that serves individuals rather than corporations, might rebound.[102]

A separate literature stakes out the poles of debate about legal tech's effect on the structure of the legal services industry.[103] Some predict that legal tech will doom BigLaw's leveraged business model by allowing smaller firms to perform as well as larger ones without the leverage—i.e., small armies of associates—that BigLaw has uniquely had at its disposal.[104] Legal tech may also reduce BigLaw's economies of scale by sharpening the case intake and risk management of smaller firms and litigation financiers.[105] Others,

---

legal tech is "tapping into a latent market of previously unserved individuals or taking business away from lawyers"). On claiming, see William L.F. Felstiner Richard L. Abel & Austin Sarat, *The Emergence and Transformation of Disputes: Naming, Blaming, Claiming . . .* , 15 LAW & SOC'Y REV. 631 (1980); DAVID M. ENGEL, THE MYTH OF THE LITIGIOUS SOCIETY: WHY WE DON'T SUE 20-36 (2016).

102 Bill Henderson, *The Decline of the PeopleLaw Sector*, LEGAL EVOLUTION (Nov. 19, 2017), https://www.legalevolution.org/2017/11/decline-peoplelaw-sector-037 [https://perma.cc/B483-KGBJ] (noting the $7 billion decline in legal services provided to individuals rather than organizations between 2007 and 2012); WILLIAM D. HENDERSON, STATE BAR OF CALIFORNIA LEGAL MARKET LANDSCAPE REPORT 17, 19 (2018), http://board.calbar.ca.gov/Agenda.aspx?id=14807&tid=0& show=100018904&s=true#10026438 [https://perma.cc/6WZ7-PAYM] (noting law's "lagging productivity" problem and arguing that legal tech can mitigate the "deteriorating economics of lawyers serving individual clients"). It is also possible, however, that legal tech tools that automate relatively simple legal tasks will hit solo practitioners and smaller firms hardest by taking away lower-level legal work. Remus & Levy, *supra* note 1, at 518-19.

103 The best overall account of the relationship between organizational forms, business models, and legal tech's proliferation is Armour & Sako, *supra* note 92.

104 Alarie et al., *supra* note 14, at 121 (predicting legal tech will leave lawyers "less tethered to working in large law firms"); Yoon, *supra* note 22, at 457 (predicting lawyers will be "less reliant on law firm economies of scale, empowering solo practitioners and small-firm lawyers"). For an overview, see Luis Garciano & Thomas N. Hubbard, *Specialization, Firms, and Markets: The Division of Labor Within and Between Law Firms*, 25 J.L. ECON. & ORG. 339 (2009). A further line of argument holds that the capital-intensive nature of technological innovation will create pressure "to move away from the traditional professional partnership model (P²) towards more managed professional businesses." Armour & Sako, *supra* note 92, at 3. Note, however, that this will require reforms relaxing legal ethical rules prohibiting fee-splitting and non-lawyer ownership of firms, both of which were adopted in the United Kingdom more than ten years ago. *Id.* But that process has only recently begun in the United States, and its future success is not yet certain. *See* David Freeman Engstrom, *Post-COVID Courts*, 68 UCLA L. REV. DISCOURSE 246 (2020) (reviewing the current state of deregulation in Utah, Arizona, and California).

105 John C. Coffee Jr., *Understanding the Plaintiff's Attorney: The Implications of Economic Theory for Private Enforcement of Law Through Class and Derivative Actions*, 86 COLUM. L. REV. 669, 706-12 (1986) (discussing the factors that explain the small size of plaintiff's firms, including the inability to adequately bring in cases that diversify their portfolio); Nora Freeman Engstrom, *Re-Re-Financing Civil Litigation: How Lawyer Lending Might Remake the American Litigation Landscape, Again*, 61 UCLA L. REV. DISCOURSE 110, 112 (2013) (noting that "changes to lawyers' financial, social, and business structures have the power to influence case outcomes"). In addition, legal tech may empower corporate clients to do their own legal work by "in-sourcing" it and thus moving it from outside counsel to general counsel offices. *See* Daniel N. Kluttz & Deirdre K. Mulligan, *Automated Decision Support Technologies and the Legal Profession*, 34 BERKELEY TECH. L.J. 853, 889 (2019) (concluding that lawyers already rely "on non-lawyer support staff and vendor judgment for a variety of tasks"); HENDERSON, *supra* note 102, at 15 (suggesting that investment in new tools will

however, have their doubts. Indeed, many legal tech tools will not be off-the-shelf tools developed by entrepreneurs and delivered across the industry, but rather more tailored, bespoke ones designed via lawyer-technologist collaborations *within* law firms. And here larger firms, with privileged access to data that comes with their repeat-player status and their ability to build internal capacity, may enjoy a decisive advantage.[106] Indeed, even the crustiest of white-shoe law firms—for instance, New York's Cravath—have built dedicated data analytics groups,[107] and others are actively entering the legal tech space and marketing their own proprietary tools.[108] While many

result in an increased pace of change). Legal tech can also leave smaller firms with more time to perform the core work of legal representation. Some studies suggest that smaller firms spend the lion's share of their time on administrative tasks and business development. *See* HENDERSON, *supra* note 102, at 14 (noting that 33% of a lawyer's workday is typically spent on business development).

106 *See* Alschner, *supra* note 78 (manuscript at 1, 10) (noting monopoly control of large legal text corpora by a set of "data haves," particularly "large legal service providers" like Westlaw and Lexis). Westlaw is defending that monopoly with litigation against smaller start-ups. Rhys Dipshan, *ROSS Rejects Westlaw Infringement Claims, Says Thomson Reuters Using Litigation as "Weapon,"* LAW.COM LEGALTECH NEWS (May 7, 2020., 12:11 PM), https://www.law.com/legaltechnews/2020/05/07/ross-rejects-westlaw-infringement-claims-says-thomson-reuters-using-litigation-as-weapon [https://perma.cc/E7YM-EPLT].

107 *See* Scott B. Reents, CRAVATH, https://www.cravath.com/sreents [https://perma.cc/QZ6A-RRWF] (noting that the head of the data analytics and e-discovery team "advises clients on defensible approaches to the preservation, collection, search and analysis of digital evidence"); *see also* Armour & Sako, *supra* note 92, at 23 ("So far, engagement with AI and related technologies has been limited to very large law firms. This tracks the economics of implementing AI solutions fixed costs and increasing returns to scale."); Josh Becker, *Legal Analytics and the Evolving Practice of Law*, LAW.COM (June 25, 2020, 11:12 AM), https://www.law.com/2020/06/25/legal-analytics-and-the-evolving-practice-of-law [https://perma.cc/B8W7-URP9] (noting similar developments at Winston & Strawn, DLA Piper); Knowledge@Wharton, *The Next Legal Challenge: Getting Law Firms to Use Analytics*, WHARTON (Nov. 22, 2019), https://knowledge.wharton.upenn.edu/article/getting-law-firms-to-use-analytics [https://perma.cc/W6T9-CG7W] (same as to Cozen O'Connor); Nicole Clark, *The Last Frontier: The Promise of AI-Powered Legal Analytics in Texas*, LAW.COM (July 13, 2020, 5:23 PM), https://www.law.com/texaslawyer/2020/07/13/the-last-frontier-the-promise-of-ai-powered-legal-analytics-in-texas [https://perma.cc/UGC2-VSWR] (same as to Ogletree Deakins); Roy Strom, *A Big Law Firm Aims to (Partly) Automate Big Law*, BLOOMBERG (Dec. 16, 2019, 5:00 AM), https://www.bloomberg.com/news/articles/2019-12-16/a-big-law-firm-aims-to-partly-automate-big-law-before-deadline [https://perma.cc/2QCX-QKBE] (same as to Wilson Sonsini); Rick Merrill, *Sheppard Mullin Embraces Cutting-Edge Legal Technology with Gavelytics Partnership*, MEDIUM (May 30, 2018), https://medium.com/@gavelytics/sheppard-mullin-embraces-cutting-edge-legal-technology-with-gavelytics-partnership-471097eec089 [https://perma.cc/QM86-TJHM] (same as to Sheppard Mullin). For still more examples and a league table of law firms and in-house counsels building internal data analytics capacity, see Kate Beioley, *Workplace Litigation: Why US Employers Are Turning to Data*, FIN. TIMES (Dec. 10, 2019), https://www.ft.com/content/865832b4-0486-11ea-a958-5e9b7282cbd1 [https://perma.cc/VQ4U-SGH6].

108 *See* Katz, *supra* note 1, at 945. Littler Mendelson provides a notable example of one of these firms. *See* Roy Strom, *Littler Mendelson Gambles on Data Mining as Competition Changes*, LAW.COM (Oct. 26, 2016, 6:21 PM), https://www.law.com/sites/almstaff/2016/10/26/littler-mendelson-gambles-on-data-mining-as-competition-changes [https://perma.cc/QNF3-Y9HP]; MARKO MRKONICH, ET AL., LITTLER MENDELSON, THE LITTLER REPORT: THE BIG MOVE TOWARD BIG DATA IN EMPLOYMENT (2015), https://www.littler.com/files/wp_big_data_8-04-15.pdf

entrepreneurs talk of disrupting the industry, legal tech may not spell doom for BigLaw. It may provide a new profit center.[109]

Both of these debates—about substitutes and complements and the structure of the legal services industry—center on profitability and so bear only a weak relationship to the question of how legal tech will reshape the litigation system or its rules. But a final strand of the debate turns toward legal tech's effect on the professional authority and orientation of lawyers and, in so doing, moves closer to procedural concerns. Summarizing a diffuse literature, two dynamics loom largest: lawyer de-skilling and lawyer de-centering. Both proceed from the premise that legal tech's rise will not merely *displace* lawyers but rather effect a subtler reshaping of relationships among lawyers, courts, and clients by introducing new kinds of professionals into litigation and by diminishing lawyers' professional agency and skill.[110]

Deskilling comes through reductions in learning opportunities and "automation bias," defined as uncritical reliance on machine outputs.[111] Both dynamics lead lawyers to invest less, and have fewer opportunities to invest,[112] in the skillsets and knowledge necessary to validate and check machine outputs.[113] The unhappy result might be a segregated profession, with tech-savvy domain experts developing and using highly effective, skill-augmenting tools, and the rest of the profession progressively losing its capacity to understand those uses or counter or question their use.[114] More generally, the

---

[https://perma.cc/QG2D-H67W]; Littler CaseSmart, https://www.littler.com/service-solutions/littler-casesmart [https://perma.cc/SE2W-5NFM]; Press Release, Littler Mendelson, Introducing Littler OnDemand: A Data-Driven Solution for Employment Law Advice and Counsel (May 9, 2019), https://www.littler.com/publication-press/press/introducing-littler-ondemand-data-driven-solution-employment-law-advice-and [https://perma.cc/53GB-L2G7].

109 This may be especially true as lawyers' professional monopoly is relaxed and non-lawyer ownership of firms expands, providing new sources of capital and making data accessibility central. *See* Armour & Sako, *supra* note 92, at 13.

110 *See* Kluttz & Mulligan, *supra* note 105, at 853 (noting quiet revolution centered not on lawyer displacement but rather a subtler reshaping of professional role and status).

111 *See* John D. Lee & Bobbie D. Seppelt, *Human Factors in Automation Design*, *in* SPRINGER HANDBOOK OF AUTOMATION 417 (Shimon Y. Nof ed., 2009) (analyzing risks from failure to address automation's restructuring of tasks, including deskilling).

112 Medicine provides a useful analogy here, particularly the rise of robotic surgery, which has reduced training opportunities and pushed surgical residents into "shadow learning" practices. *See* Matthew Beane, *Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail*, 64 ADMIN. SCI. Q. 87, 87 (2019) (defining shadow learning as "an interconnected set of norm- and policy-changing practices enacted extensively, opportunistically, and in relative isolation that allowed only a minority of robotic surgical trainees to come to competence").

113 Hildebrandt, *supra* note 87, at 156 (recommending that scrutinizing "the 'intestines' of the data mining process" should become part of a lawyer's training).

114 Kluttz & Mulligan, *supra* note 105, at 884 (suggesting that most lawyers currently do not understand "testing and validation terms and metrics"); Pasquale & Cashwell, *supra* note 90, at 81 (recommending that all lawyers should have access to this technology if it becomes significant enough to impact advocacy).

legal profession could experience a kind of "judgmental atrophy" or a creeping "epistemic sclerosis."[115] We return to this idea shortly in considering legal tech's effect on rule of law.

Decentering is easiest to see in the e-discovery context. As use of TAR proliferates, discovery disputes will play out as expert battles in which dueling technologists opine about the propriety of data manipulations, modeling choices, and performance metrics.[116] TAR thus encroaches on the legal profession's control over one of the fundamental domains of litigation procedure and, as commentators have put it, "transform[s] litigation procedure—traditionally the exclusive domain of judges and lawyers—into a domain that is shared with computer scientists, commercial vendors, and others."[117] Put another way, lawyers will progressively cede professional jurisdiction to technologists. Even if law remains a profession with most of its current trappings—partial professional monopoly, self-regulation, sizeable returns to talent—the result will be a steady leakage of professional status and authority.[118]

2. Legal Tech and Rule of Law

A second strand of an emerging literature explores legal tech's implications for conceptions and implementations of rule of law. Some of the most dramatic follow from the futurist predictions noted previously about a state of "legal singularity" and a "self-driving" legal order.[119] Those scenarios, however, hold few implications for litigation over the near- to medium-term—and, if realized, will ultimately render much of the legal tech toolkit irrelevant because procedures, judicial discretion, and legal systems as we know them will cease to exist.[120]

A more tractable set of rule-of-law concerns posed by legal tech's advance can be bucketed into two categories: *personnel*-based concerns (i.e., concerns rooted in the changing role and status of lawyers) and *process*-based concerns (i.e., concerns rooted in coming changes to the process of adjudicating legal claims).

---

115 *See* Hildebrandt, *supra* note 30, at 32-33.

116 *See* Remus, *supra* note 34, at 1711 (explaining that attorneys' lack of technological expertise now requires the use of experts to defend their use of certain technologies).

117 *Id.* at 1710-11 (warning that lawyers are "ceding control" over procedure); *see also* Kluttz & Mulligan, supra note 105, at 889 (reporting that lawyers increasingly rely upon "non-lawyer support staff and vendor judgment" on selection and configuration of systems and model-testing); Shannon H. Kitzer, *Garbage In, Garbage Out: Is Seed Set Disclosure a Necessary Check on Technology-Assisted Review and Should Courts Require Disclosure?*, 1 J.L. TECH. & POL'Y 197, 201 (2018) (same).

118 *See, e.g.*, Armour & Sako, *supra* note 92, at 10 (noting the emergence of a new and contested "expert division of labor" within the legal system); Pasquale, *supra* note 3, at 5 (arguing that an automated legal system "shifts personal responsibility from attorneys, regulators, and judges, to those coding their would-be replacements").

119 *See supra* notes 9–13 and accompanying text.

120 *See, e.g.*, Casey & Niblett, *supra* note 10, at 1436, 1440.

Understanding each highlights some trade-offs of legal tech's incorporation into the civil justice system that may demand a procedural response.

Personnel-based concerns extend from the twin processes of lawyer "decentering" and "deskilling" just noted.[121] Hildebrandt puts it well: As the practice of law is progressively turned over to technologists, there are fewer "legal natives" with a "vested interest in or experience with the issues of the Rule of Law."[122] In lawyers' stead will come an array of non-lawyer experts with a very different worldview, built around using and promoting technology. As one commentator puts it, these new legal professionals "have no reason to recognize, much less incorporate within their opinions, lawyers' ethical obligations to clients, the courts, and the public"—and, worse, may have "internalized their employers' profit motive."[123]

Process-based concerns, in contrast, stem from the basic insight that legal tech tools will actively shape law rather than just being used to deploy it. A fast-growing literature on legal tech offers a master class in jurisprudence, from Holmes to Hart to Hayek and from Langdell to Long Fuller to Llewelyn (and back again), taking on the efficiency of the common law, rules versus standards, dialogic versus more instrumental approaches to law, and even conceptions of legality itself. This is neither the time nor place to review its many tributaries. But much of this thinking comes back to a single key insight: Outcome prediction tools, to choose one part of the legal tech toolkit, are not, as Frank Pasquale and Glyn Cashwell artfully put it, just "a camera trained on the judicial system," but rather an "engine of influence."[124] One easily glimpsed possibility that follows is that outcome prediction tools, and likely other parts of the legal tech toolkit as well, will progressively drain the system of its flexibility, its adaptive capacity, and its dialogic core. Legal automation, on this view, brings "a fast and refined prediction of the relevant legal effect"[125] and thus achieves one of the highest (but by no means only) purposes of law: fast and cheap resolution of disputes. But it comes at a steep

---

[121] Personnel-based concerns might also include legal tech's effects on judicial self-conception and the psychology and practice of judging. For a recent and insightful study of the long-term effect of Israel's Legal-Net system, see generally Amnon Reichman, Yair Sagy & Shlomi Balaban, *From a Panacea to a Panopticon: The Use and Misuse of Technology in the Regulation of Judges*, 71 HASTINGS L.J. 589 (2020). *See also* Orna Rabinovich-Einy & Ethan Katsh, *The New New Courts*, 67 AM. U. L. REV. 165, 165 (2017) (analyzing the implications of alternative and online dispute resolution on the judicial system). In particular, any effort to quantify the production of legal decisions will likely skew thinking toward efficiency and away from other values (quality, dignity of litigants, etc.).

[122] Hildebrandt, *supra* note 87, at 149 (noting how lawyers buttress rule of law); Hildebrandt, *supra* note 30, at 12 (same). For an historical account, see DANIEL R. ERNST, TOCQUEVILLE'S NIGHTMARE: THE ADMINISTRATIVE STATE EMERGES IN AMERICA, 1900-1940 (2014).

[123] Remus, *supra* note 34, at 1711 (describing the danger of deferring to non-lawyer IT experts).

[124] *See* Pasquale & Cashwell, *supra* note 90, at 67 (describing the influence of the deployment of predictive models on the judicial system).

[125] Hildebrandt, *supra* note 30, at 21 (describing positive impacts of artificial intelligence on litigation).

cost, draining the law of its capacity to adapt to new developments or to ventilate legal rules in formal, public interpretive exercises.[126] At the extreme, legal tech may even work a change in our conception of legality itself, substituting prediction for persuasion and reason-giving and shifting law's normative center to a Skinnerian model of cognition in which law is merely "a black-boxed transformation of inputs into outputs."[127] The resulting "reductionism and functionalism," and the related elevation of predictability over vitality, does more than impair the system's adaptive capacity. The system also loses its legitimacy as a way to manage social conflict when the process of enforcing collective value judgments plays out in server farms rather than a messy deliberative and adjudicatory process, even where machine predictions prove perfectly accurate.[128]

3. Legal Tech and Distribution

A third and final broad implication of legal tech, related to but distinct from the other two, is political in the classic distributive sense of that term—the "who

---

[126] *Id.* at 21-23; Re & Solow-Niederman, *supra* note 3, at 253-54 (arguing that legal tech will favor "codified" justice over "equitable justice" and thus "standardization" over "discretion"); McGinnis & Wasick, *supra* note 22, at 1046 (exploring computation's implications for the persistent tension in law between "comprehensibility" and "predictability"). In Re and Solow-Niederman's view, legal tech will favor codified justice, which privileges efficiency and uniformity, over equitable justice, which is more discretionary, contextual, and dynamic. Lost in the process will be values of mercy, mitigation, and extenuation. Worse, the move to codified justice will have a "self-legitimating power" because datification will center and foreground values linked to available data, that are themselves conducive to further automation. *See* Re & Solow-Niederman, *supra* note 3, at 270 (likening this to the man who searches for lost keys under the streetlamp); *see also* Kleinberg et al., *supra* note 31, at 494 (noting data's tendency to orient organizations toward questions that can be quantified and computed); Harry Surden, *The Variable Determinacy Thesis*, 12 COLUM. SCI. & TECH. L. REV. 1, 8 (2011) (arguing that AI might generate pressure for legislators and rulemakers to adapt law in ways that are susceptible of computational analysis). But this is far from clear: Some predict a collapse of rules and standards—or, rather, standards into rules. Casey & Niblett, *supra* note 10, at 1405. Powerful machines using advanced analytics on rich datasets can take account of more fine-grained differences across cases than a human judge, not fewer. This suggests that there may be a temporal dynamic that AI-driven adjudication must push through, something that Re and Solow-Niederman wisely acknowledge. Re & Solow-Niederman, *supra* note 3, at 260 ("At least in the near term, . . . AI adjudication will not embody equitable justice."). In other words, with technological advances, one could code equity into AI adjudicators or legal tech tools, and so AI adjudication may ultimately prove more *perfectible*. *Id.* at 268.

[127] Pasquale & Cashwell, *supra* note 90, at 65 (describing how the use of algorithmic predictive analytics in judicial contexts is an "emerging jurisprudence of behaviorism"); *see also* Sheppard, *supra* note 11, at 36.

[128] *See* Joshua P. Davis, *Law Without Mind: AI, Ethics, and Jurisprudence*, 55 CAL. W. L. REV. 165 (2018) (providing reasons to not rely on technology for legal issues due to the inability of machines to make moral judgements); Mireille Hildebrandt, *Law as Information in the Era of Data-Driven Agency*, 79 MOD. L. REV. 1, 7 (2016) (noting AI's threat of "mindless agency"). *See generally* LON L. FULLER, THE MORALITY OF LAW (1964) (describing the connections between legal morality and major themes in law).

gets what, when and how" of political science[129] or, in Marc Galanter's litigation-specific formulation, whether and how the "haves" come out ahead.[130]

Legal tech's promise is that if, as some predict, new tech tools erode BigLaw's economies of scale and empowers smaller firms and solo practitioners, then one might expect a leveling of the playing field between "haves" and "have nots."[131] Perhaps most important of all, and to circle back to a claim explored previously, supply and demand are endogenous to legal costs.[132] The declining cost of supplying legal services may render claims marketable that cannot currently draw counsel, particularly given the paring back of aggregation mechanisms like the class action.[133] For champions of civil rights or consumer protection, among others, the result could be a golden age of litigation in which those priced out of the current litigation system can more reliably vindicate their rights.[134] Legal tech, on this view, might have its greatest impact in areas where would-be litigants with quality claims are not currently being served.[135]

If some see legal tech as a democratizing force, others have their doubts. A common theme is that legal tech will at best replicate and at worst exacerbate existing power and resource disparities within the litigation system. As already noted, few legal tech tools are turnkey; most require significant mid-stream customization in order to enrich search results, refine predictions of case outcomes, or iteratively label documents for relevance and privilege.[136] As a result, legal tech tools may merely replicate asymmetries in the quantity and quality of lawyering within the system. Similar dynamics might play out as the process of lawyer decentering noted previously steadily converts traditional procedural wrangling, particularly around discovery, into

---

[129] HAROLD D. LASSWELL, POLITICS: WHO GETS WHAT, WHEN AND HOW (1990).

[130] *See generally* Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC'Y REV. 95 (1974) (explaining why repeat-players and those with more resources and power come out ahead in litigation).

[131] *See* Paul Gowder, *Transformative Legal Technology and the Rule of Law*, 68 U. TORONTO L.J. 82, 83 (2018); Yoon, *supra* note 22, at 457, 470 (describing increased lawyer productivity and benefits to both lawyers and clients as a result of emerging technologies).

[132] *See supra* note 101 and accompanying text.

[133] For example, Radvocate (now FairShake) is an internet-based tool that helps individual consumers pursue arbitration claims, in return for a contingency fee. *See* FAIRSHAKE, https://fairshake.com [https://perma.cc/245V-P6KP].

[134] *See supra* notes 101–105 and accompanying text.

[135] Other examples: Firms are developing Q/A systems covering aspects of tax or privacy law compliance that are unlikely to justify retaining a lawyer. *See Australian Privacy Compliance Packages*, NORTON ROSE FULBRIGHT (Dec. 2018), https://www.nortonrosefulbright.com/en/knowledge/publications/bcf33bd9/australian-privacy-compliance-packages [https://perma.cc/XC64-3N7A] (displaying examples of packages clients can buy to remain in compliance with new Australian regulations).

[136] *See supra* notes 75–81 and accompanying text.

expert battles between technologists.[137] In the current system, better-heeled litigants can afford better experts and so may systematically win out over less-resourced ones. Legal tech may reproduce or amplify those effects.

Darker predictions imagine a world that is more different in kind than degree. As just noted, some of the best legal tech tools may emerge from BigLaw's in-house expertise and privileged data access, something larger law firms are more likely to have.[138] From there, one can imagine a more significant divergence in the counsel available to the better and worse off, with the "haves" enjoying the services of a new kind of super-lawyer whose superior skill and connections are further augmented by software, and the "have nots" settling for unboosted human lawyering or, perhaps worse, an inferior machine-only version.[139] Bleaker still is the possibility that legal tech may yield proportionally greater deployment of law by "haves" than "have nots."[140] Witness, for instance, the use of robo-approaches in evictions, mortgage foreclosures, or consumer credit disputes,[141] or more recent reports that law firms, tech companies, and large employers and retailers (e.g., Walmart) are using advanced analytics to draft responsive pleadings and discovery requests at low cost and make outcome predictions that can guide their settlement calculus and litigation strategy.[142]

---

[137] Endo, *supra* note 34, at 862-68 (discussing the normative trade-off between economic efficiency and participation); Remus, *supra* note 34, at 1711 (detailing the increased reliance on non-lawyer IT experts with the increased use of technology in legal matters).

[138] *See supra* notes 106–109 and accompanying text.

[139] *See* Remus & Levy, *supra* note 1, at 551 (imagining a similar "two-tiered system"); *see also* Michael Livermore & Dan Rockmore, *France Kicks Data Scientists Out of Its Courts*, SLATE (June 21, 2019, 7:30 AM), https://slate.com/technology/2019/06/france-has-banned-judicial-analytics-to-analyze-the-courts.html [https://perma.cc/C3LV-V35C]. Livermore and Rockmore elaborate on these potential inequities:

> Already, access to legal services is doled out according to ability to pay, with money buying higher-quality representation. A.I. could supercharge this phenomenon, with only the rich able to buy the latest software, while the rest of us are stuck with wetware humans with their limited memory and processing speed. Alternatively, government cutbacks in legal services for the poor might eventually result in over-reliance on subpar A.I. tools, with (possibly) more nimble human lawyers and customized software reserved for the well-heeled.

Id.

[140] *See* Gowder, *supra* note 131, at 90 (arguing that legal tech may expand legal services against subordinated groups, worsening the relative situation between the "haves" and "have nots").

[141] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 4 (2014) (discussing effects of using algorithms to the detriment of individuals).

[142] Though the precise bounds of the practice are hard to determine, there is substantial evidence that BigLaw firms (among them Ogletree Deakins and Littler Mendelson) and legal tech companies (among them LegalMation) are working with Walmart and other large entities to develop automated tools that automatically draft responsive pleadings and discovery requests in high-volume litigation areas such as employment and personal injury ("slip and fall" cases), saving defendants valuable time, and also predict case outcomes in those cases. *See, e.g.*, *Episode 33: Using AI in Litigation—Thomas Suh (LegalMation*

On this view, it may be better-heeled litigants—who are also litigation's repeat-players[143]—who will capture the benefits of procedural streamlining and data analytics. Rather than leveling the playing field, legal tech may make it easier for employers, creditors, and landlords to prosecute cases against employees, debtors, and tenants.[144]

Finally, legal tech's distributive effects will turn on the economic and legal structure of access to it. Legal tech tools that act as force-multipliers cannot democratize litigation if most lawyers and clients are priced out of their use. Nor can they serve as levelers if the falloff from the advanced versions sitting behind paywalls and simpler open-source versions is too steep.[145] For those who worry above all about legal tech's distributive effects, the overriding imperative going forward will be to ensure that all parties to disputes have access to key technology and the data necessary to power it.[146] But that access, and legal tech's broader incorporation into the litigation system, will also be modulated by legal structures, including the twin workings of IP and trade secrets evidentiary privilege and legal-ethical rules that govern non-lawyer firm ownership. The legal structure of access to legal tech and the types of firm organization and business models that are permitted and prohibited under legal-ethical rules will profoundly shape its cost structure and its development path, yielding wide access to legal tech's fruits or, to the

---

Co-Founder), TECHNICALLY LEGAL PODCAST (May 27, 2020), https://tlpodcast.com/episode-33-using-ai-in-litigation-thomas-suh-legalmation-co-founder [https://perma.cc/P6ZW-JGGA]; Press Release, Ogletree Deakins, Ogletree Deakins and LegalMation Announce Innovative Partnership (Jan. 9, 2019), https://ogletree.com/media-center/press-releases/2019-01-09/ogletree-deakins-and-legalmation-announce-innovative-partnership [https://perma.cc/D99P-NY96]. For a PowerPoint touting a Walmart/Ogletree/LegalMation tool, see Alan Bryan, Senior Assoc. Gen. Couns. Legal Operations, Walmart, Patrick DiDomenico, Chief Knowledge Officer, Ogletree Deakins, Tariq Abdullah, Senior Dir. Legal Operations and Data & Analytics, Walmart, & James Lee, Chief Exec. Officer, LegalMation, Using A.I. to Digitize Lawsuits to Perform Actionable Data Analytics, Presentation at the 2019 Corporate Legal Operations Consortium Vegas Institute (May 15, 2019), [https://perma.cc/B4C2-XY3K]. For a concerned analysis, see Patricia Barnes, *Artificial Intelligence Further Exacerbates Inequality in Discrimination Lawsuits*, FORBES (Aug. 26, 2019, 5:40 PM), https://www.forbes.com/sites/patriciagbarnes/2019/08/26/artificial-intelligence-further-exacerbates-inequality-in-discrimination-lawsuits [https://perma.cc/5WBU-XXVW].

   143   *See* Galanter, *supra* note 130, at 107-114 (detailing how rules favor repeat players rather than one-shotters).

   144   *See* Pasquale & Cashwell, *supra* note 90, at 65 (stating that predictive analytics will be used by "richer litigants to gain advantages over poorer ones").

   145   Brian Sheppard, *Why Digitizing Harvard's Law Library May Not Improve Access to Justice*, BLOOMBERG L. (Nov. 12, 2015, 2:21 PM), https://news.bloomberglaw.com/business-and-practice/why-digitizing-harvards-law-library-may-not-improve-access-to-justice [https://perma.cc/77V9-T3VR] (noting that Ravel Law makes its baseline tools freely available but its more advanced analytics and research sit behind a paywall).

   146   *See* Pasquale & Cashwell, *supra* note 90, at 81 (arguing that jurisdictions should develop rules to "level the playing field"). On data accessibility, see Re & Solow-Niederman, *supra* note 3, at 285 (calling for "public option" legal tech and data accessibility as "an institutional counterweight to proprietary datasets").

contrary, ensuring that legal tech remains a proprietary tool of litigation's "haves."[147] We return to these ideas below, particularly in Section II.C, because civil procedure rules, particularly the work product doctrine, may act to bolster or curtail those rights.

<p style="text-align:center">*   *   *</p>

More than a hundred years ago, Justice Holmes, in *The Path of the Law*, wrote: "For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics."[148] Shorn of its fusty, turn-of-the-century diction, this statement could just as easily have come, in 2021 rather than 1897, from the mouth of a legal tech entrepreneur. And there is for sure an element of puffery in such claims, both then and now, in light of NLP's significant technical challenges. There is, however, no doubt that substantial change is afoot, even if its particulars remain fuzzy. In the next two decades, we will likely see a substantial change in how lawyers do their work, sometimes for the better, sometimes not. We are also likely to see a diminution in lawyers' professional role, stature, and authority, but also a newly powerful cadre of tech-savvy super-lawyers. And we will witness a shift in the litigation landscape toward both democratization and domination.

But amidst all of this contingency is a single, undeniable certainty: Over the near to medium-term, legal tech will be shaped in important part by how the litigation system and, in particular, judges armed with little more than the rules of civil procedure manage and guide its uptake. In the next Part, we aim to add concreteness to current thinking about legal tech by asking, in three case studies, how particular legal tech tools will reshape the litigation system and how civil procedure can, or should, adapt in response.

II. Legal Tech and Civil Procedure: Three Case Studies

This Part climbs down from the heights of thinking about legal tech's longer-run effects on law and the legal system and offers three concrete cuts at legal tech's evolution over the near- to medium-term. As to each, we ask: Assuming

---

147 *See* Remus, *supra* note 34, at 1715 ("[P]atent protection threatens to increase unequal access to predictive-coding technologies, which will entrench existing disparities in resources and power."); Armour & Sako, *supra* note 92 (interrogating the relationship between lawyer regulation, available business models and firm structures, and legal tech innovation); Henderson, *supra* note 102, at 20-24 (investigating how ethics rules regulate the market for legal services and legal tech); Gillian K. Hadfield, *The Cost of Law: Promoting Access to Justice Through the (Un)corporate Practice of Law*, 38 Int'l Rev. L. & Econ. 43 (2014) (discussing how access to justice issues arise under the existing infrastructure for provision of legal services).

148 Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 Harv. L. Rev. 457, 469 (1897).

continuing advances in legal tech tools over the next ten or fifteen years, how will legal tech change litigation, and how might civil procedure adapt?

This posture, we believe, brings two advantages. First, in focusing on the near- to medium-term, we aim to avoid some of the pitfalls of working at the intersection of law and technology. Technology can evolve in wholly unexpected ways, and even short-range predictions about technological innovation can be deeply misguided.[149] Yet hewing too closely to present-day technology can yield an analysis akin to rearranging deck chairs on the Titanic, imagining a modest set of altered litigation realities, and a set of legal procedural responses, just before a wave of innovation upends the system. Limiting our inquiry to the foreseeable trajectory of legal tech over the near- to medium-term aims to steer between these two extremes.

Second, by grounding our analysis in civil procedure, we gain traction by focusing in on a discrete set of more litigation-centered legal tech tools. We further maintain focus by building each case study around specific legal tech tools, a prediction about their effect on the distribution of costs or information within the system, and potential amendments to one or more specific civil procedural rules or doctrines. More specifically, Section II.A links TAR tools, the distribution of litigation costs within the system, and rules governing proportionality and pleading. Section II.B links outcome-prediction tools, the distribution of information as between judges and litigants, and the rules and doctrines that govern forum-shopping. Section II.C. links TAR and various legal analytics and outcome-prediction tools and the distribution of information among litigants to the work product doctrine.

A. *Predictive Coding, Proportionality, and Plausibility Pleading*

No analysis of legal tech, civil procedure, and the future of the adversarial system would be complete without attending to the technological revolution in discovery practices that is already well underway. This section glimpses the new world of discovery as new technological tools proliferate and then spins out the implications for civil procedure, focusing in particular on legal tech's capacity to shift the distribution of litigation costs within the system.

---

149 *See* Ryan Calo, *Commuting to Mars: A Response to Professors Abraham and Rabin*, 105 VA. L. REV. ONLINE 84, 88-90 (2019) (noting that technological change occurs against the backdrop of social, cultural, and economic forces that shape the trajectory of the technology itself).

1. The New World of Discovery

Discovery is variously described as the "heart," "backbone," "focal point," and "foundation" of American litigation, and with good reason.[150] The essential purpose of discovery, after all, is to identify material facts that prove or disprove a claim. Moreover, because discovery is the backdrop for everything that follows it—motions, trial, appeal—cases are often won or lost at the discovery stage. And indeed, at least since the 1938 Federal Rules of Civil Procedure, discovery has been deliberately structured—some say overly so—to facilitate settlement and thus obviate the need for trial at all.[151] To be sure, this was not always so. In an era of "non-suits," trial *was* discovery and, if new facts surfaced, a do-over called. Even today, large swathes of cases—low-stakes auto accidents, among many others—involve no discovery at all.[152] Still, no part of the litigation system has generated more heated debate in

---

150 Diego A. Zambrano, *Discovery as Regulation*, 119 MICH. L. REV. 71, 72 (2020); Nora Freeman Engstrom, *The Lessons of Lone Pine*, 129 YALE L.J. 2, 67 (2019); Robin Page West, *Letters for Litigators*, 31 LITIGATION, Spring 2005, at 21, 25.

151 J. Maria Glover, *The Federal Rules of Civil Settlement*, 87 N.Y.U. L. REV. 1713, 1752 (2012) (describing the settlement-focused features of the system but noting that "[i]n the context of modern litigation, the notion that more discovery will always promote better merits-based resolution of claims needs to be revisited"); John H. Langbein, *The Disappearance of Civil Trial in the United States*, 122 YALE L.J. 522, 526, 542 (2012) (describing that discovery has had the effect of displacing trial in most cases); Stephen C. Yeazell, *The Misunderstood Consequences of Modern Civil Process*, 1994 WIS. L. REV. 631, 647-48, 671 (1994) (asserting that the 1938 Federal Rules of Civil Procedure empowered trial judges to urge parties to settle). Compared to civil law countries and even most common law ones, American discovery yields far broader information exchange. *See generally* Owen M. Fiss, Comment, *Against Settlement*, 93 YALE L.J. 1073 (1984) (critiquing the system's orientation toward settlement); Judith Resnik, *Managerial Judges*, 96 HARV. L. REV. 374, 378 (1982) (noting that the creation of pretrial discovery rights facilitated the rise of managerial judging).

152 The data is noisy. PAUL R. CONNOLLY, EDITH A. HOLLEMAN & MICHAEL J. KUHLMAN, FED. JUDICIAL CTR., JUDICIAL CONTROLS AND THE CIVIL LITIGATIVE PROCESS: DISCOVERY 28 (1978) (suggesting more than half of cases involve little discovery); EMERY G. LEE III & THOMAS E. WILLGING, FED. JUDICIAL CTR., NATIONAL CASEBASED CIVIL RULES SURVEY: PRELIMINARY REPORT TO THE JUDICIAL CONFERENCE ADVISORY COMMITTEE ON CIVIL RULES 8 (2009) (finding no discovery in approximately fifteen percent of cases); James S. Kakalik, Deborah R. Hensler, Daniel McCaffrey, Marian Oshio, Nicholas M. Pace & Mary E. Viana, *Discovery Management: Further Analysis of the Civil Justice Reform Act Evaluation Data*, 39 B.C. L. REV. 613, 636 (1998) (reporting that thirty-eight percent of general civil cases do not involve lawyer hours worked on discovery); Susan Keilitz, Roger A. Hanson & Henry W.K. Daley, *Is Civil Discovery in State Trial Courts out of Control?*, STATE CT. J., Spring 1993, at 9 (noting that forty-two percent of general litigation cases lack discovery, and thirty-seven percent of those with discovery had three or fewer pieces of discovery).

recent decades, including a parade of reform proposals[153] and frequent amendments to the federal and state rules.[154]

Discovery is a lightning rod not just because of its centrality in modern litigation, but also because it has been one of the most dynamic parts of the system. Two seismic developments have remade the discovery process in recent decades. The first is the pervasive digitization of society, which has fueled a steady rise, beginning in the 1990s, of electronically stored information, or ESI. Some estimate that the total amount of digitized material in the current "Big Data" era doubles every few years—a kind of Moore's law of digital information.[155] The second development is the advent of new automated tools for identifying, retrieving, processing, and analyzing this crush of materials. At first, managing it meant a move beyond "linear manual review," in which lawyers put eyeballs on every document, to more automated approaches centered around processing techniques that make documents machine-readable (e.g., OCR) and searchable (e.g., keywords).[156]

---

[153] *See* Seth Katsuya Endo, *Discovery Hydraulics*, 52 U.C. DAVIS L. REV. 1317, 1343-50 (2019) (summarizing reform proposals). *See also* ROBERT G. BONE, CIVIL PROCEDURE: THE ECONOMICS OF CIVIL PROCEDURE 216-19 (2003) (considering two major complaints about discovery costs: excessive discovery and abusive discovery); Robert D. Cooter & Daniel L. Rubinfeld, *An Economic Model of Legal Discovery*, 23 J. LEGAL STUD. 435, 455 (1994) (proposing a "two-part rule" to reform discovery and shift a portion of costs to the requesting party); Frank H. Easterbrook, *Discovery as Abuse*, 69 B.U. L. REV. 635, 642, 646, 648 (1989) (investigating several proposals, such as "break[ing] the process [] into smaller chunks," to reform discovery); Martin H. Redish, *Electronic Discovery and the Litigation Matrix*, 51 DUKE L.J. 561, 608-14 (2001) (proposing a reformed cost-shifting model for electronic discovery).

[154] *See* Stephen B. Burbank, Sean Farhang & Herbert M. Kritzer, *Private Enforcement*, 17 LEWIS & CLARK L. REV. 637, 657 n. 79 (2013) (noting amendments made to the Federal Rules in response to complaints about discovery); Endo, *supra* note 153, at 1343-50 (describing a variety of discovery reforms that have been proposed and implemented since 2006).

[155] *See* JOHN GANTZ & DAVID REINSEL, THE DIGITAL UNIVERSE IN 2020: BIG DATA, BIGGER DIGITAL SHADOWS, AND BIGGEST GROWTH IN THE FAR EAST (2012) ("From [2012] until 2020, the digital universe will about double every two years."); PETER LYMAN & HAL R. VARIAN, HOW MUCH INFORMATION? (2003), https://groups.ischool.berkeley.edu/archive/how-much-info-2003 [https://perma.cc/7ZGS-V869] ("We estimate that the amount of new information stored on paper, film, magnetic, and optical media has about doubled in the last three years."). For discussion in the legal context, see Jason R. Baron, *Law in the Age of Exabytes: Some Further Thoughts on 'Information Inflation' and Current Issues in E-Discovery Search*, 17 RICH. J.L. & TECH. 1, 4-5 (2011).

[156] *See, e.g.*, Remus & Levy, *supra* note 1, at 515 (describing the evolution of document review, which at first relied on "deductive instructions to search documents for keywords" and more recently is shifting to predictive coding); *see also* Christian, *supra* note 34, at 496-97, 524 (describing keyword searches and noting that the progression to TAR stems in part from the unsuitability of keyword searches for truly massive document productions). *See generally* Symposium, *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189 (2007). Precise estimates of the pervasiveness of automated forms of discovery are hard to come by. But one recent survey of federal judges found that some forty-five percent of district judges and fifty-nine percent of magistrate judges had ordered use of a computer search methodology of one sort or another for discovery of voluminous materials, especially ESI.

More recently, automated discovery has leaped ahead with the advent of the TAR (or "predictive coding") tools described previously that use machine learning classifiers to flag relevant and privileged documents. Taken together, these two trends—proliferating ESI and new automated ways of analyzing it—have progressively remade the world of discovery and ensured that the discovery process, already a burning topic, has remained at the white-hot center of debate about procedure and litigation's role in American society.

This new world of discovery has generated a predictable set of concerns, some new and some reaching back to the old world. The first is the acceleration of the trend away from comprehensiveness in discovery.[157] At the creation of the Federal Rules of Civil Procedure in the 1930s, some cheered that the new rules permitted "an almost unlimited discovery."[158] Soon after, Justice Murphy penned an iconic statement of comprehensiveness: "Mutual knowledge of *all* the relevant facts gathered by both parties is essential to proper litigation."[159] But that utopian ideal has steadily eroded as litigation has grown in scale and complexity. The first dents in the armor came with the 1976 Pound Conference, which some see as the wellhead of a cost-obsessed, anti-litigation strain that has defined American law and politics ever since.[160] A rule-based version came in 1983, when the Judicial Conference amended Rule 26 of the federal rules to require proportionality between discovery requests and the needs of a case, and then again in 2006, when the rules were amended to adapt e-discovery to this goal. The crush of ESI and increasing use of automation has been the final nail in the coffin. Indeed, in complex litigations, it is already the case that the *Hickman* mindset of exhaustive discovery has been eclipsed by a far less ambitious approach in which discovery is more a negotiation about quantitative error tolerance—that is, the production of an acceptable percentage of documents at an acceptable level of accuracy—and a truly exhaustive surfacing of evidence is only rarely cost-justified.[161]

See Paul W. Grimm, *Are We Insane? The Quest for Proportionality in the Discovery Rules of the Federal Rules of Civil Procedure*, 36 REV. LITIG. 117, 138 tbl.3 (2017).

[157] Remus, *supra* note 34, at 1718 (highlighting the trend towards proportionality, which entails "agreement on the production of a particular percentage of documents at a particular level of accuracy . . . .").

[158] Edson R. Sunderland, *Discovery Before Trial Under the New Federal Rules*, 15 TENN. L. REV. 737, 738 (1939).

[159] Hickman v. Taylor, 329 U.S. 495 (1947) (emphasis added).

[160] STEPHEN B. BURBANK & SEAN FARHANG, RIGHTS AND RETRENCHMENT: THE COUNTERREVOLUTION AGAINST FEDERAL LITIGATION 101 (2017) ("Some scholars have 'characterized the . . . Pound Conference . . . as the most important event in the counteroffensive against notice pleading and broad discovery.'").

[161] Remus, *supra* note 34, at 1718 (defining proportionality within a predictive-coding regime); Geoffrey P. Miller, *On the Costs of Civil Justice*, 80 TEX. L. REV. 2115, 2117-18 (2002) (finding that the American litigation system is "at an inefficient point" when comparing trade-offs between the costs of error and the costs of procedure). Despite the commitment of these resources, no one could or

A second broad concern, and one noted previously, is that the rising sophistication of e-discovery tools is eroding lawyers' professional jurisdiction and authority.[162] But TAR does not cut lawyers out of the equation entirely. Rather, they perform traditional document review on a subset of a production to create a "labeled" set of documents—or a "seed set"—to train the model, then engage in further such efforts as the system iteratively moves toward a best model.[163] But beyond this lawyer-centered data-labeling role, TAR is a highly technical exercise. It involves an array of methodological choices, as evidenced by a growing literature evaluating seed set selection strategies,[164] choices among "learning protocols" at the more iterative stage of model training,[165] and performance metrics,[166] that sit far beyond the average lawyer's ken. Lawyers, the worry goes, will progressively cede professional authority to technologists (the people who develop, tune, and deploy the models) and technologist experts (the people who opine about the quality of this or that approach before judges in motions practice) in a key procedural domain.

should expect perfection from this process. A representative contemporary judicial statement of lowered ambition comes in *Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc.*, No. 11-6189, 2014 WL 584300, at *2 (S.D.N.Y. Feb. 14, 2014) ("The production of documents in litigation such as this is a herculean undertaking, requiring an army of personnel and the production of an extraordinary volume of documents . . . . Despite the commitment of these resources, no one could or should expect perfection from this process.").

162 *See supra* notes 110–118, and accompanying text.

163 *See supra* notes 34–35, and accompanying text.

164 *See* Christian J. Mahoney, Nathaniel Huber-Fliflet, Haozhen Zhao, Jianping Zhang, Peter Gronvall & Shi Ye, *Evaluation of Seed Set Selection Approaches and Active Learning Strategies in Predictive Coding*, 1 PROC. INT'L WORKSHOP ON A.I. AND INTELLIGENT ASSISTANCE FOR LEGAL PROS. DIGIT. WORKPLACE 2 (2019) (evaluating seed set selection strategies, including "random" sampling and more complex, "judgmental" sampling in which attorneys use case-specific knowledge).

165 At the second iterative step, one chooses an "active learning protocol" to select further training documents to add to the seed set. Protocol options include: the strongest matches, borderline matches, a random set of matches, or a mix of each. Mahoney et al., *supra* note 164, at 3; *see also* Gordon V. Cormack & Maura R. Grossman, Autonomy and Reliability of Continuous Active Learning for Technology-Assisted Review 4 (Apr. 26, 2015) (unpublished manuscript), https://arxiv.org/pdf/1504.06868.pdf [https://perma.cc/83TH-TM5V] (discussing active learning in the context of TAR); Rishi Chhatwal, Nathaniel Huuber-Fliflet, Robert Keeling, Jianping Zhang & Haozhen Zhao, *Empirical Evaluations of Active Learning Strategies in Legal Document Review*, 2017 IEEE INT'L CONF. ON BIG DATA 1428 (providing an empirical analysis of the utility of active learning in the legal domain).

166 These metrics include measures of when the system has stabilized and requires no further iteration. They also include performance metrics, including recall (the percentage of targeted documents returned by the algorithm); precision (the percentage of recalled documents that meet targeting criteria); and the F1-Score (the harmonic mean of precision and recall—i.e., $2*(P*R)/(P+R)$). *See* ASHLEY, *supra* note 14, at 114. A final metric is the area under the receiver operating characteristic curve ("AUC"), which plots true positive and false positive rates against a set of possible thresholds and thus gives information for various levels of precision and recall how confident one can be that the model captures relevant documents. *Id.* at 257. An AUC score of 100% is perfect; a score of 50% means it is no more likely than chance that all relevant documents have been ranked higher than irrelevant ones. *Id.*

Perhaps the most significant concern connecting the old and new worlds of discovery is litigation costs. The American approach to liberal discovery, embedded in an adversarial scheme in which parties (mostly) bear their own costs, has many virtues, but it creates two glaring incentive problems. First, the propounding party can externalize a large share of the cost of discovery requests onto her adversary, constrained only by the cost the party subsequently incurs in requesting and then processing and reviewing the material produced.[167] Second, tasking the responding party with responsibility for determining relevance creates "cross-party agency costs."[168] The responding party both has superior information about the value of the discovery in question and is also strongly incentivized to produce as little relevant material as possible. By producing limited *relevant* information, she can minimize her own discovery costs and legal exposure and, by burying that information in a mountain of extraneous materials, raise her adversary's review costs or even prevent her adversary from finding the "needle in the haystack."[169] The system in effect trusts a party to act as her "adversary's agent" and "decide whether a document is useful to her adversary's case."[170]

The result of these misaligned incentives is two types of cost concern. One is that excessive discovery can yield high aggregate litigation costs relative to case stakes—i.e., the (dis)proportionality concern that has long occupied courts and rulemakers. Adjudication of these disputes, the argument goes, diverts valuable social resources that could be better deployed elsewhere. The other is that discovery costs are often asymmetrically distributed among the parties. As a result, one side in a litigation, often the defendant, bears more of the cost of litigation, giving the other side, often the plaintiff, undue settlement leverage and perhaps even yielding settlements in cases lacking any merit at all.[171] These two types of litigation

---

167 *See* Jonah B. Gelbach & Bruce H. Kobayashi, *The Law and Economics of Proportionality in Discovery*, 50 GA. L. REV. 1093, 1103 (2016) (discussing the externalization of costs in modern discovery practice).

168 *Id.* at 1104 (arguing that the responding party's attorney sometimes has much discretion in deciding whether a document is useful to the adversary's case, thus forcing the responding party's attorney to act as his adversary's agent).

169 *See* Bruce H. Kobayashi, *Law's Information Revolution as Procedural Reform: Predictive Search as a Solution to the In Terrorem Effect of Externalized Discovery Costs*, 5 U. ILL. L. REV. 1473, 1478, 1498-1501 (2014) (examining the "nature and effect of cross-party agency costs" in the context of discovery); *see also* Gelbach & Kobayashi, *supra* note 167, at 1105 (noting that, when a responding party limits its effort in accurately sorting between relevant and non-relevant documents, that party increases its adversary's cost of discovery while simultaneously reducing its own).

170 Gelbach & Kobayashi, *supra* note 167, at 1104; *see also* Kobayashi, *supra* note 169, at 1478.

171 *See* Martin H. Redish & Colleen McNamara, *Back to the Future: Discovery Cost Allocation and Modern Procedural Theory*, 79 GEO. WASH. L. REV. 775 (2011) (challenging the long-held assumption that the producing party bears the costs of discovery, arguing that there are important differences between discovery costs in production and the normal costs in preparing a legal defense);

costs need not, of course, yield an inefficient system. Even costly litigations or litigations featuring wide cost asymmetries produce a mix of social costs (excessive litigation) and social benefits (deterrence, surfacing additional wrongdoing) that can be complex depending on the circumstances.[172] In economics terms, the private and the social motive can diverge in ways that lead to either too much or too little litigation.

Despite their centrality in legal and policy debates, discovery costs have drawn little careful rigorous empirical inquiry, particularly recently.[173] Indeed, much of the best empirical evidence dates back to the 1970s and 1980s, as private litigation grew in importance as a regulatory mechanism and concern about litigation's inefficiencies crested.[174] While differences across litigation systems (federal, state) and litigation types make generalizations difficult, three core findings establish an empirical baseline about cost concerns while leaving plenty of room for debate as to their salience. First, discovery is a substantial, though probably not dominant, source of litigation costs—perhaps one-quarter to one-third of the total.[175] But discovery costs also

BONE, *supra* note 153, at 45-50 (discussing the asymmetrical discovery costs through a hypothetical that focuses on information asymmetry between parties); Cooter & Rubinfeld, *supra* note 153, at 437 (noting that asymmetrical transaction costs in discovery distort the terms of settlement, giving one party undue advantage over the other); David Rosenberg & Steven Shavell, *A Model in Which Suits Are Brought for Their Nuisance Value*, 5 INT'L REV. L. & ECON. 3, 4-5 (1985) (providing numerical examples to illustrate that, even in instances where the plaintiff brings an unmeritorious claim, settlement might be the most cost-effective option for the defendant).

172 *See* Gelbach & Kobayashi, *supra* note 167, at 1102 ("Since external costs are associated with too much litigation, while external social benefits are associated with too little, there are *gross* effects operating in both directions. As a matter of simple arithmetic, then, the *net* impact of these gross effects might point in either direction."); *see also* Steven Shavell, *The Fundamental Divergence Between the Private and the Social Motive to Use the Legal System*, 26 J. LEGAL STUD. 575, 575 (1997) (arguing that litigants do not think about negative or positive externalities—which include cost and deterrence respectively—resulting in litigation that is either "socially excessive" or "inadequate").

173 *See, e.g.*, Alexandra D. Lahav, *A Proposal to End Discovery Abuse*, 71 VAND. L. REV. 2037, 2052 (2018) (lamenting lack of empiricism and calling for required docketing of discovery requests); *see also* Judith A. McKenna & Elizabeth C. Wiggins, *Empirical Research on Civil Discovery*, 39 B.C. L. REV. 785, 796-97 (1998) (making similar complaints about lack of empiricism surrounding the costs of modern discovery).

174 On the anti-litigation trend, see BURBANK & FARHANG, *supra* note 160; Resnik, *supra* note 151. Key empirical studies throughout the 1970s and 1980s include CONNOLLY ET AL., *supra* note 152; WILLIAM A. GLASER, PRETRIAL DISCOVERY AND THE ADVERSARY SYSTEM 162-77 (1968), Wayne D. Brazil, *Civil Discovery: Lawyers' Views of Its Effectiveness, Its Principal Problems and Abuses*, 1980 AM. BAR FOUND. RSCH. J. 787, Wayne D. Brazil, *Views from the Front Lines: Observations by Chicago Lawyers About the System of Civil Discovery*, 1980 AM. BAR FOUND. RES. J. 217; Susan Keilitz, Roger Hanson & Richard Semiatin, *Attorneys' Views of Civil Discovery*, JUDGES' J., Spring 1993, at 2; Keilitz et al., *supra* note 152; David M. Trubek, Austin Sarat, William L.F. Felstiner, Herbert M. Kritzer & Joel B. Grossman, *The Costs of Ordinary Litigation*, 31 UCLA L. REV. 72 (1983).

175 *See* LEE & WILLGING, *supra* note 152, at 38-39 (2009) (considering individual federal cases in 2008, including those that went to trial, and reporting that the median portion of total litigation costs incurred for discovery was 20% for plaintiffs and twenty-seven percent for defendants); Kakalik et. al., *supra* note 152, at 637 (finding that lawyer hours per litigant was 232 hours, with an average of

appear to vary significantly across cases.[176] Second, while many lawyers believe that discovery costs are often out of proportion with case stakes,[177] the best recent study puts attorney estimates of discovery's proportion at 1.6% to 3.3% of total stakes in the median case, and roughly one-quarter to one-third of total stakes at the top end.[178] Third, litigation cost asymmetries are real but vary in magnitude throughout the system. For instance, older studies found that, in patent cases, discovery consumed more than twice the defendant's costs as it did the plaintiff's.[179] More recent data suggests that, in the more general run of cases, defendant-side discovery costs are somewhat smaller—perhaps not quite double plaintiffs'.[180]

---

83 hours, or thirty-six percent spent on discovery, including motions practice); Trubek et al., *supra* note 174, at 90-91 (finding that only 16.7% of attorney time was spent on discovery, setting aside outlier "megacases"); Paula Hannaford-Agor, *Measuring the Cost of Civil Litigation: Findings from a Survey of Trial Lawyers*, VOIR DIRE, Spring 2013, at 22, 26 (reporting survey results and concluding that in six common types of disputes discovery accounted for twenty-two percent of lawyer and paralegal hours). Older studies tend to find relatively higher discovery costs. *See* Thomas E. Willging, Donna Stienstra, John Shapard & Dean Miletich, *An Empirical Study of Discovery and Disclosure Practice Under the 1993 Federal Rule Amendments*, 39 B.C. L. REV. 525, 531 (1998) (finding that, among cases involving discovery, fifty percent of litigation costs came in discovery); *Judicial Conference Adopts Rules Changes, Confronts Projected Budget Shortfalls*, U.S. CTS. (Sept. 15, 1999), https://www.uscourts.gov/news/1999/09/15/judicial-conference-adopts-rules-changes-confronts-projected-budget-shortfalls [https://perma.cc/KMR6-BJDY] (reporting a similar figure).

176 *See* LEE & WILLGING, *supra* note 152, at 38-39 (reporting attorney-based estimates of the proportion of total litigation costs consumed by discovery as low as 0.1% and as high as eighty percent across roughly 1,000 cases); NICHOLAS M. PACE & LAURA ZAKARAS, RAND INST. FOR CIV. JUST., WHERE THE MONEY GOES: UNDERSTANDING LITIGANT EXPENDITURES FOR PRODUCING ELECTRONIC DISCOVERY 17-18 (2012), https://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1208.pdf [https://perma.cc/KEA6-UXDR] (reporting highly variable total discovery costs across various types of cases).

177 *See* LEE & WILLGING, *supra* note 152, at 28 (reporting that attorneys in 25% of cases believed discovery costs are too high relative to AIC). In addition, studies report that most lawyers erroneously believe that discovery consumes about two-thirds of litigation costs and estimate that 50% would be a more appropriate number. A.B.A., ABA SECTION OF LITIGATION MEMBER SURVEY ON CIVIL PRACTICE: DETAILED REPORT 98 (2009); REBECCA M. HAMBURG & MATTHEW C. KOSKI, NAT'L EMP. LAWS. ASS'N, SUMMARY OF RESULTS OF FEDERAL JUDICIAL CENTER SURVEY OF NELA MEMBERS 34 (2010).

178	LEE & WILLGING, *supra* note 152, at 42-43.

179	GLASER, *supra* note 174, at 166.

180	The most comprehensive recent federal-level study is the 2009 Federal Judicial Center report, which finds that the median proportion of total litigation costs incurred in discovery was twenty percent of $15,000 in total litigation costs, or $3,000, for plaintiffs, and twenty-seven percent of $20,000 in total litigation costs, or $5,400, for defendants; these costs rose by between five to ten percentage points for both parties in cases with ESI. *See* LEE & WILLGING, *supra* note 152, at 2. By contrast, a 1998 study, which found that fifty percent of total litigation costs went to discovery, showed no difference across plaintiffs and defendants. *See* Willging et al., *supra* note 175, at 531. It is important to remember that cost statistics are necessarily based on cases selected for litigation. *See infra* notes 252–253 and accompanying text (exploring theories of selection of disputes for litigation).

An even harder question to interrogate empirically is where costs will go in the new world of AI-boosted discovery.[181] One oft-articulated view is that discovery costs will continue to rise because of the ever-growing universe of discoverable material—"infinite ESI"—and because ever cheaper digital storage will allow us to keep all of it.[182] Some of this has come from predictable precincts—the Chamber of Commerce and other anti-litigation standard bearers who have worked hard, and successfully, to establish an often-misleading "cost-and-delay narrative" about litigation.[183] But it has also come from more official and less conflicted quarters, including judges[184] and rulemakers,[185] among others.[186]

However, close attention to data, economic theory, and a technical understanding of TAR and related e-discovery tools, suggests something very nearly the opposite may prove true. Indeed, largely missing from the debate is a key and, we believe, unmistakable observation: In recent decades digitization has produced a substantial uptick in the volume of ESI, while the advanced analytics necessary to manage that volume have lagged behind; as TAR continues to proliferate and improve, however, the discovery cost curve

---

181 *See* Kluttz & Mulligan, *supra* note 105 at 854 ("[P]redictive coding methods . . . challenge the current model for evaluating whether and how tools are appropriate for legal practice."); NORTON ROSE FULBRIGHT, 2018 LITIGATION TRENDS ANNUAL SURVEY: PERSPECTIVES FROM CORPORATE COUNSEL 3 (2018) (reporting a survey of corporate counsel finding that fifty-four percent of companies had used TAR).

182 *See, e.g.*, Andrew Jay Peck, *Foreword*, 26 REGENT U. L. REV. 1, 3 (2014) (noting the explosion of digital information and inability of discovery methods to catch up); *see also* John H. Beisner, *Discovering a Better Way: The Need for Effective Civil Litigation Reform*, 60 DUKE L.J. 547, 550 (2010) (arguing that "exponential growth" in electronic documents has fueled "abusive discovery"); Endo, *supra* note 153, at 1338 (noting that cost concerns animate most discovery reforms).

183 Danya Shocair Reda, *The Cost-and-Delay Narrative in Civil Justice Reform: Its Fallacies and Functions*, 90 OR. L. REV. 1085, 1085-86 (2012) (reviewing and criticizing the debate). For an example of claims from within the interest group landscape, see AM. COLL. OF TRIAL LAWYERS & INST. FOR THE ADVANCEMENT OF THE AM. LEGAL SYS., FINAL REPORT ON THE JOINT PROJECT OF THE AMERICAN COLLEGE OF TRIAL LAWYERS TASK FORCE ON DISCOVERY AND THE INSTITUTE FOR THE ADVANCEMENT OF THE AMERICAN LEGAL SYSTEM 2 (2009) [hereinafter COLL. OF TRIAL LAWYERS] (concluding that there are serious problems in the civil justice system and that the discovery structure is in need of reform).

184 *See* Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 311 (S.D.N.Y. 2003) (referring to an "exponentially" increasing universe of discovery material).

185 *See* Memorandum from David G. Campbell, Chair, Advisory Comm. on Civ. Rules, on Report of Advisory Committee on Civil Rules to Jeffery S. Sutton, Chair, Comm. on Rules of Prac. & Proc. 24 (May 2, 2014) (noting impact of the "information explosion" on the 1993 amendments to Rule 26).

186 *See, e.g.*, PACE & ZAKARAS, *supra* note 176, at 26 (explaining participant's concerns with "high costs" of discovery review); *The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production*, 19 SEDONA CONF. J. 1, 59 (2018) (describing the growth of ESI).

is likely to bend down more quickly than the digitization curve bends up.[187] This, we submit, will have important implications for procedure, and may drain the proportionality constraints built into federal and state civil procedure rules of much of their importance. Further, powerful new e-discovery tools seem poised to steadily narrow the litigation cost asymmetries that have motivated a second key procedural reform in recent years: *Twombly/Iqbal*'s shift in the pleading rules.

2. Proportionality's Retreat in a Frictionless World

Growing concern about litigation costs has spurred a wide catalog of reform ideas in recent decades,[188] but the reform that judges and policymakers have arguably preferred above all others is the imposition of proportionality constraints on discovery. As noted previously, proportionality became part of the federal rules in 1983, but it was beefed up significantly in 2006, when the Advisory Committee added an ESI-specific rule designed to guard against "undue burden and cost."[189] In 2015, the Committee re-centered the proportionality constraint by moving it front and center in Rule 26(b)(1)'s provisions governing discovery's scope, although the operative text changed little.[190] In its current guise, Rule 26(b)(1) now permits discovery

> regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case, considering the importance of the issues at stake in the action, 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.[191]

Many states have followed suit.[192]

The newly centered proportionality provisions have, by most accounts, had a substantial effect, drawing both criticism and praise.[193] However, there are

---

[187] Against the backdrop of a chorus of voices focused on ever-increasing costs, only a few commentators have acknowledged this possibility. *See, e.g.*, Grimm, *supra* note 156, at 167 (noting that new predictive technologies may reduce the cost of ESI discovery).

[188] *See supra* notes 153–154 and accompanying text.

[189] FED. R. CIV. P. 26(b)(2)(B).

[190] The proportionality mandate was moved from FED. R. CIV. P. 26(b)(2)(C)(iii) to FED. R. CIV. P. 26(b)(1) in 2015.

[191] FED. R. CIV. P. 26(b)(1).

[192] *See, e.g.*, ARIZ. R. CIV. P. 26(b)(1); COLO. R. CIV. P. 26(b)(1); ILL. SUP. CT. R. 201(c)(3); MD. CODE ANN., CT. R. § 2-402(b)(1); KAN. STAT. ANN. § 60-226(b)(1) (2019); MINN. R. CIV. P. 1 & 26.02(b)(2)-(3); OKLA. STAT. tit. 12 § 3226(B)(1)(a) (2020); UTAH R. CIV. P. 26(b)(2); VT. R. CIV. P. 26(b)(1); WYO. R. CIV. P. 26(b)(1)).

[193] *But see* Richard L. Marcus, *Discovery Containment Redux*, 39 B.C. L. REV. 747, 773-74 (1998) (reviewing claims but arguing that the shift in the rules has not had a major impact); *see also* Ion

two reasons to believe that legal tech will shift the ground out from under proportionality constraints, progressively eliminating much of their force.[194] First, there are reasons to doubt the pervasive claims about "infinite" ESI that have helped drive reform efforts. In trial litigation, much discovery comes from communication, which is, in important ways, bounded by the limits of human attention and cognition and may bear only a weak relationship to the growth in other kinds of digital materials. There are limits to the quantity of email that even large, sprawling organizations can generate.[195]

Second, and more importantly, evidence is mounting that continued diffusion of TAR tools will reduce, perhaps substantially, total discovery costs. Only recently, commentators expressed doubt about TAR's accuracy and efficiency relative to manual review.[196] But a growing cluster of studies establishes that well-implemented TAR tools are as good as, and often better than, purely human review in terms of recall (i.e., the proportion of documents in the total pool of documents that the tool accurately identifies as relevant) and almost certainly *better* than humans in precision (i.e., the proportion of documents among those the tool identifies that are in fact relevant).[197] And, they

Meyn, *The Haves of Procedure*, 60 WM. & MARY L. REV. 1765, 1791 (2019) (decrying test's vagueness and resulting judicial discretion); Gelbach & Kobayashi, *supra* note 167, at 1117 (observing that proportionality brings "subjectivity and a reduction of predictability").

194 *Cf.* Endo, *supra* note 153, at 1354-55 (providing a framework for considering proportionality issues in discovery); *see also* Ralph C. Losey, *Predictive Coding and the Proportionality Doctrine: A Marriage Made in Big Data*, 26 REGENT U. L. REV. 7, 15-16 (2013) (arguing that predictive coding is the answer to the proportionality doctrine); Peck, *supra* note 182, at 3 (suggesting that technology offers solutions to the discovery problems it created).

195 *Cf.* THE RADICATI GRP., EMAIL STATISTICS REPORT, 2015–2019 at 3-4 (2015) (estimating that the number of business emails per user per day will grow from 122 to 126 from 2015-2019, an increase of just three percent over 4 years).

196 *See* Remus, *supra* note 34, at 1707 (noting a general lack of validation of existing tools and likely variation in their quality).

197 "Well-implemented" is the key qualifier here and includes the quality of the algorithm itself, the technologists who deploy it, the data set, and the broader workflow and pipeline around each of these things. Importantly, the legal tech industry's marketing efforts, and even a growing academic literature, frequently overstate TAR's capabilities. Perhaps the best source of empirical studies of TAR's efficacy come from the Legal Track Interactive Task Studies at the Text Retrieval Conference (TREC), convened by the National Institute of Standards and Technology (NIST) from 2008 to 2010. *See generally TREC Legal Track*, https://trec-legal.umiacs.umd.edu [https://perma.cc/SLM3-8GLB]. In what were, in effect, competitions among invited commercial and academic entrants, few achieved an F_1 statistic higher than fifty percent, and fewer still achieved an F_1 higher than seventy percent. For an explanation of F_1, see supra note 166. All told, only a small number of entrants were conclusively better than human, eyes-on review. This is important, for many of the most frequently cited claims as to TAR's efficacy are based on the TREC results. *See, e.g.*, Grossman & Cormack, *supra* note 14, at 2-5 (offering evidence of TAR's efficiency and accuracy based on an analysis of data collected from the TREC). For an accounting of academic studies and conference exercises up through 2012, see PACE & ZAKARAS, *supra* note 175, at 61-69. Of course, it is possible that the technology has improved since the TREC studies, which are nearly ten years old. More recent studies report impressive findings. *See, e.g.*, Chhatwal et al., *supra* note 165, at 1433 (reporting ninety percent

do so at a fraction of the cost.[198] Put another way, well-implemented TAR may not consistently capture substantially more relevant or privileged documents, but it yields less surplusage and requires a fraction of attorney time.[199]

All of this comes with caveats—and spotlights future avenues for research. First, the performance metrics that underpin claims about TAR's cost-savings are not ironclad. We lack perfectly accurate "ground truth" because we can never "know" which documents in a production are relevant within Rule 26's meaning because that meaning is subjective and contestable.[200] However, while it is possible that skepticism about these measures will slow TAR's advance, the better bet remains that it will continue to improve and earn judicial sanction.[201] Second, it is important to concede that TAR, while

---

recall based on review of only forty percent of the documents and concluding automation can be better than linear human review). *But see* Robert Keeling, Rishi Chhatwal, Peter Gronvall & Nathaniel Huber-Fliflet, *Humans Against Machines: Reaffirming the Superiority of Human Attorneys in Legal Document Review and Examining the Limitations of Algorithmic Approaches to Discovery*, 26 RICH. J.L. & TECH., no. 3, 2020, at 2 (challenging the "prevailing wisdom" about TAR's efficacy relative to manual review and critiquing studies). A new round of highly-controlled TREC-like studies might provide a more reliable sense of current performance.

198 Grossman & Cormack, *supra* note 14, at 43 (suggesting "a fifty-fold savings over exhaustive manual review"). Some courts have begun to hear testimony on TAR's efficiency gains. *See* Dynamo Holdings Ltd. P'ship v. Comm'r, 143 T.C. 183, 194 (2014) (featuring expert testimony that TAR would reduce discovery costs from around $500k to around $80k). Courts are also increasingly willing to tout TAR's efficiencies. *See* Chevron Corp. v. Donzinger, No. 11-0691, 2013 WL 1087236, at *33 n.255 (S.D.N.Y. Mar. 15, 2013) (rejecting argument about burdensome discovery and noting that TAR allows for review at a "fraction of the cost of human reviewers"); Hyles v. New York City, No. 10-3119, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016) (stating that "TAR is the best and most efficient search tool" for discovery).

199 TAR will not be the only source of cost-savings. Cheaper storage means that fewer resources will also be spent extracting ESI from inefficient storage (e.g., backup tapes). *See* SEDONA CONFERENCE WORKING GRP. ON ELEC. DOCUMENT RETENTION & PROD., THE SEDONA PRINCIPLES: BEST PRACTICES RECOMMENDATIONS AND PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION 5 (2d ed. 2007) ("[T]he fact that many forms of electronically stored information and media can be searched quickly and accurately by automated methods provides new efficiencies and economies.").

200 *See* Peter Bailey, Nick Craswell, Ian Soboroff, Paul Thomas, Arjen P. de Vries & Emine Yilmaz, *Relevance Assessment: Are Judges Exchangeable and Does It Matter?*, 31 PROC. ANNUAL INT'L ACM SIGIR CONF. ON RES. & DEV. INFO. RETRIEVAL 667 (2008) (suggesting that judges vary significantly in their assessments of what material is relevant). The best studies manage this concern by reporting numerous pair-wise comparisons of multiple manual (human) reviews and machine outputs. *See* PACE & ZAKARAS, *supra* note 176, at 61-66 (summarizing the methods and results of four studies comparing TAR to manual review). The TREC studies create a single measure of relevance by appointing a "Topic Authority" who played the role of a senior attorney directing a discovery effort by developing relevance criteria, and who was accessible to participants. The resulting standard thus did not aspire to intersubjective validity but instead provided a single, albeit subjective, standard of relevance for purposes of judging performance.

201 The general trend is toward judicial acceptance. *See, e.g.*, Moore v. Publicis Groupe, 287 F.R.D. 182, 187 (S.D.N.Y. 2012) (concluding that TAR "works better than most of the alternatives, if not all of the [present] alternatives"), *adopted sub nom*. Moore v. Publicis Groupe SA, No. 11-1279, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012); Nat'l Day Laborer Org. Network v. U.S. Immigr. &

reducing the need for lawyers, will add new entries to the cost side of the discovery ledger, including software, technologists, and litigation experts.[202] These new inputs could render TAR inefficient in smaller-scale productions.[203] The more general question is whether the uptick in costs will be less or more than the downtick in lawyer time necessary to perform linear manual review. The smart money is on less, but one cannot predict with certainty where reality will land.

A final caveat looms larger: Studies proclaiming TAR's superiority assume a static litigation system. This assumption, however, may not hold. To begin, cost reductions can reshape how much of the task is demanded and/or supplied, and this is no less true in law than elsewhere.[204] The unit cost of discovery can drop precipitously, but aggregate costs may not budge if judges proceed to green-light ever more expansive discovery requests. Moreover, most studies touting TAR do not account for possible shifts under the discovery rules. But in a growing set of cases, lower courts are struggling with whether to compel party cooperation.[205] In early cases, courts declined to mandate disclosure of seed sets because the parties had arrived at arrangements themselves.[206] Where conflicts arise, however, courts must

---

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*, 143 T.C. at 192 ("[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) (holding that TAR is "an acceptable way to search for relevant ESI in appropriate cases"). *But see* Progressive Casualty Ins. Co. v. Delaney, No. 11-00678, 2014 WL 3563467, at *11 (D. Nev. 2014) (refusing party request to use TAR but mostly because of the party's bad faith in abandoning an agreed-to protocol). For cases in Anglo-American legal systems beyond the U.S., including the U.K., Ireland, and Canada, see Christian, *supra* note 34, at 503-10.

202 Remus, *supra* note 34, at 1707 (noting that existing studies fail to account for the possibility of these other costs).

203 Remus & Levy, *supra* note 1, at 534 (claiming that training and review, in particular, might render TAR inefficient for small classifications); *see also* Endo, *supra* note 34, at 855 (assessing whether predictive coding is fiscally efficient across all cases).

204 *See supra* notes 101–102 and accompanying text. Basic economic reasoning suggests the point: if technology reduces the cost of producing a good or service, that means it increases supply, but if it also increases the quality of the good or service, demand also may increase. For example, improved knee replacement devices might lead to cheaper treatment of knee problems as well as more demand for that kind of treatment. With reduced cost per treated patient but more patients treated, total health spending might rise or fall as a result of the new technology. *See also* John G. Heyburn II & Francis E. McGovern, *Evaluating and Improving the MDL Process*, 38 LITIGATION 26, 28-30 (2012) (noting the dynamic relationship of supply and demand in managing MDL dockets).

205 For general law review commentary, see Kitzer, *supra* note 117, at 206; Christian, *supra* note 34, at 524-25; Tonia Hap Murphy, *Mandating Use of Predictive Coding in Electronic Discovery: An Ill-Advised Judicial Intrusion*, 50 AM. BUS. L.J. 609, 632 (2013).

206 *See Rio Tinto*, 306 F.R.D. at 129 (approving TAR protocol agreed to by parties); Bridgestone Americas, Inc. v. Int'l Bus. Mach. Corp., No. 13-1196, 2014 WL 4923014, at *1 (M.D. Tenn. July 22,

choose, with some strongly encouraging disclosure but not mandating it[207] and others requiring disclosure or making it a condition of a responding party's use of TAR.[208] Academic commentators go furthest of all, proposing that the *requesting* party be made solely responsible for constructing the seed set and tuning the machine learning model.[209] We provide a fuller analysis of the implications of these positions in Section II.C below, including the possibility that the work product doctrine might protect seed sets from disclosure. For now, it is enough to observe that each of these positions could have a range of as-yet-unanalyzed effects on the distribution of discovery costs, including perhaps *increasing* costs in certain cases.[210]

These are important caveats, and yet the broader conclusion seems sound. Short of substantial changes to current discovery rules, the near- to medium-term is likely to see a reduction in overall discovery costs. As a corollary, the proportionality concerns that have animated much recent litigation reform activity are likely to fade in importance, particularly in cases whose major costs are driven by large-corpus electronic document discovery.

2014) (same); *In re* Actos (Pioglitazone) Prod. Liab. Litig., No. 11-2299, 2012 WL 7861249, at *4-5 (W.D. La. July 27, 2012) (same).

207 *See, e.g.*, Aurora Coop. Elevator Co. v. Aventine Renewable Energy-Aurora W., LLC, No. 12-230, 2015 WL 10550240, at *2 (D. Neb. Jan. 6, 2015) (noting that Rule 26(b)(1) does not authorize discovery of the *irrelevant* documents in a seed set but then "encourag[ing]" the parties to work cooperatively); *In re* Biomet M2a Magnum Hip Implant Prod. Liab. Litig., No. 12-2391, 2013 WL 6405156, at *2 (N.D. Ind. Aug. 21, 2013) (noting lack of authority under Rule 26(b)(1) but "urg[ing] Biomet to re-think its refusal"); *Winfield v. City of New York*, No. 15-05236, 2017 WL 5664852, at *12 (S.D.N.Y. Nov. 27, 2017) ("encourag[ing]" but not requiring disclosure).

208 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"). In *Progressive*, the court refused a request to use TAR because the party advocating it had refused to share seed sets and other methodological details. Progressive Casualty Ins. Co. v. Delaney, No. 11-00678, 2014 WL 3563467, at *11 (D. Nev. 2014) ("Progressive is unwilling to engage in the type of cooperation and transparency that its own e-discovery consultant has so comprehensibly and persuasively explained is needed for a predictive coding protocol to be accepted by the court or opposing counsel as a reasonable method to search for and produce responsive ESI."). In at least one other case, the judge ordered seed set disclosure in a ruling from the bench. *Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc.*, No. 11-6189, 2014 WL 584300 (S.D.N.Y. 2014).

209 *See* Kobayashi, *supra* note 169, at 1504 (arguing that the requesting party should be responsible for the up-front costs of search).

210 As an example, privilege (as opposed to relevance) determinations cannot be re-allocated to the requesting party under the current work product doctrine and attorney-client privilege. As a result, even if the requesting party is given sole responsibility for tagging the seed set for relevance, the responding party must *still* review documents for privilege prior to turning over the seed set, embroiling *both* parties in substantial review work. Particularly in smaller-scale cases where predictive coding brings only small efficiency gains, an approach that imposes review obligations on both parties could *increase* total litigation costs.

3. Re-Centering *Twombly* and *Iqbal*

If proportionality has created a slow burn of reform skirmishes in recent decades, then recent changes to the pleading rules, anchored by the U.S. Supreme Court's opinions in *Twombly* and *Iqbal*, were more of a surprise revolution.[211] In its *Twombly* and *Iqbal* opinions, the Supreme Court, ostensibly interpreting Rule 8 of the Federal Rules of Civil Procedure, swept away the "notice pleading" system that had prevailed since the creation of the federal rules in 1938 and replaced it with a regime in which a plaintiff's complaint must assert a "plausible" claim for relief in order to withstand a motion to dismiss.[212] Heated debate has ensued about whether this is in fact just a probability requirement in fancy clothes, and lower courts have often struggled with how to implement the Court's new mandate in any other way.[213] A long academic literature of varying rigor and sophistication has also questioned whether and to what extent the change matters, particularly for specific case types (e.g., civil rights).[214] And whether or not *Twombly* and *Iqbal* have had tangible effects on *judicial* decisions, the new pleading regime may still have impacted pleading practice—for instance, causing many plaintiffs to make costly investments in pre-filing investigation to avoid dismissal.[215]

At the normative core of the *Twombly/Iqbal* debate is a value judgment about the collision of two kinds of asymmetries: asymmetric discovery costs

---

[211] *See* Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 COLUM. L. REV. 433, 492-93 (1986) (predicting the demise of notice pleading). For the cases, see *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 (2007); *Ashcroft v. Iqbal*, 556 U.S. 662 (2009).

[212] *Twombly*, 550 U.S. 544; *Iqbal*, 556 U.S. 662.

[213] *See, e.g.*, Swanson v. Citibank, N.A., 614 F.3d 400, 404 (7th Cir. 2010) ("As we understand it, the Court is saying instead that the plaintiff must give enough details about the subject-matter of the case to present a story that holds together. In other words, the court will ask itself *could* these things have happened, not *did* they happen."); Braden v. Wal-Mart Stores, Inc., 588 F.3d 585, 594 (8th Cir. 2009) ("The plausibility standard requires a plaintiff to show at the pleading stage that success on the merits is more than a 'sheer possibility.' It is not, however, a 'probability requirement.'" (internal citations omitted)).

[214] David Freeman Engstrom, *The* Twiqbal *Puzzle and Empirical Study of Civil Procedure*, 65 STAN. L. REV. 1204-06 (2013) (reviewing studies that analyze the systemic impact of *Twombly* and *Iqbal*).

[215] *See* Stephen B. Burbank, *Pleading and the Dilemmas of Modern American Procedure*, 93 JUDICATURE 109, 118 (2009) ("Perhaps the most troublesome possible consequence of Twombly and Iqbal is that it will deny access to court to plaintiffs and prospective plaintiffs with meritorious claims who cannot satisfy their requirements . . . because they lack the resources to engage in extensive pre-filing investigation or because of informational asymmetries."). There is no direct evidence of any such effect. However, studies suggest that the decisions affected other aspects of pleading practice and so it is eminently plausible. *See* Christina L. Boyd, David A. Hoffman, Zoran Obradovic & Kosta Ristovki, *Building a Taxonomy of Causes of Action in Federal Complaints*, 10 J. EMP. LEG. STUDS. 253, 274 (2013) (offering preliminary evidence that plaintiffs pled fewer causes of action after *Twombly*). *But see* REBECCA M. HAMBURG & MATTHEW C. KOSKI, SUMMARY OF RESULTS OF FEDERAL JUDICIAL CENTER SURVEY OF NELA MEMBERS, FALL 2009 (2010) (reporting survey results in which employment discrimination plaintiffs' counsel reported no change in the amount of detail in their pleadings).

and asymmetric information. The first of these, we just noted, arises from the misaligned incentives of the American system of discovery in which costs lie where they fall, allowing parties to externalize the costs of discovery requests onto adversaries. By subjecting a party's claims to pre-discovery scrutiny, *Twombly/Iqbal*'s pleading rule seeks to blunt such *in terrorem* effects on settlement.[216] But whatever its value in paring back litigation cost asymmetries, plausibility screening also creates a countervailing concern founded upon information asymmetries. Simply put, not all claimants have access to needed evidence at the pleading stage, and only coercive discovery and compulsory process can dislodge privately held information about wrongdoing. The result is that the effects of litigation cost asymmetries can be mitigated only by exacerbating information asymmetries, and vice versa. *Twombly/Iqbal*'s plausibility pleading standard is merely a choice, and a highly subjective one, along a spectrum of possible accommodations of the two concerns.

Twombly/Iqbal's balancing act may involve distributional considerations, but it is not necessarily intractable. One option is to relax the "plausibility" mandate in the subset of cases most afflicted by asymmetric information (though doing so might strain the American commitment to transubstantivity).[217] Another partial solution is phased discovery, akin to jurisdictional discovery, to target key evidentiary issues—the "jugular" of a case—at the pleading stage in order to test plausibility, but leaving the bulk of discovery to later stages, once a motion to dismiss has been beaten back.

For those cases in which document discovery is a key cost driver, TAR adds a third potential solution to this menu of options. A small, but growing, academic literature has begun to explore this possibility, and the reasoning, pivoting off of the earlier discussion of TAR's effect on proportionality, should now be familiar.[218] The core of the argument is that TAR will substantially narrow asymmetric discovery costs because a prime source of those asymmetries—review of documents for relevance and privilege—is the

---

216 *Twombly*, 550 U.S. at 559 (justifying plausibility pleading on need "to avoid the potentially enormous expense of discovery in cases with no 'reasonably founded hope that the [discovery] process will reveal relevant evidence' to support a . . . claim"); *see also* Samuel Issacharoff & Geoffrey Miller, *An Information-Forcing Approach to the Motion to Dismiss*, 5 J. LEGAL ANALYSIS 437, 448 (2013) (suggesting that heightened pleading standards serve to "prevent[] deadweight losses through fruitless discovery"); Jonah B. Gelbach, Note, *Locking the Doors to Discovery? Assessing the Effects of Twombly and Iqbal on Access to Discovery*, 121 YALE L.J. 2270, 2277 (2012) (showing that *Twombly/Iqbal* standards prevent discovery in some cases that would otherwise reach it).

217 *See, e.g.*, *Swanson*, 614 F.3d at 404-05 (discussing the application of the *Twombly/Iqbal* standard to different substantive areas of law); *see also* Andrew Blair-Stanek, *Twombly is the Logical Extension of the* Mathews v. Eldridge *Test to Discovery*, 62 FLA. L. REV. 1, 38 (2010) (discussing the ill effects of *Twombly* and *Iqbal* on civil rights and discrimination cases).

218 Kobayashi, *supra* note 169, at 1502.

discovery cost that is most directly abated by TAR.[219] Moreover, these review costs tend to be unevenly distributed between requesting and responding parties, the argument continues, because the responding party must review the full set of collected documents for both relevance and privilege before producing them while the requesting party receives and reviews only the distilled set.[220]

As with our claims around proportionality, the empirical case for TAR's capacity to narrow cost asymmetries is less than ironclad. Time will tell, and will open significant opportunities for future research. First, TAR's capacity to mitigate the *in terrorem* effect of cost asymmetries may not be evenly felt on both sides of the "v," yielding substantial reductions in discovery costs among responding and requesting parties alike. Requesting parties, for instance, might utilize TAR to more efficiently distill a large document production for review. If new algorithmic tools cut the requesting party's costs as much or nearly as much as the responding party's, then cost asymmetries, and the *in terrorem* effect they underwrite, may not budge. TAR's capacity to narrow cost asymmetries may also be limited in smaller-stakes cases. Because TAR's economies fade as the quantity of discovery declines, there is a point at which the fixed cost of software, seed set construction, and model tuning is not worth the candle. This is important, because at least some empirical evidence suggests that cost asymmetries may be at their widest in smaller-stakes cases, not the mega-litigations that feature most prominently in litigation reform debates.[221]

Second, TAR's capacity to mitigate the effects of cost asymmetries will—as with TAR's effect on proportionality—depend on how courts modulate its use by litigants. As just noted, lower courts are grappling with how much inter-party cooperation to require when implementing TAR protocols, and some academics go further and advocate a shift to a "task allocation" rule in which the requesting party performs the work, and bears the cost, of constructing the seed set and training the model as a way to limit cost externalization and cross-party agency costs.[222] However, privilege determinations, we also noted, may be non-delegable, making any re-allocation of discovery tasks to the requesting party at best partial.[223] The result is that TAR might narrow, but likely cannot flatten, discovery costs.

These and other objections to TAR's capacity to mitigate litigation cost concerns provide fruitful avenues for future research, both theoretical and

---

[219] PACE & ZAKARAS, *supra* note 176, at 41 (estimating that maximum savings can be achieved by increasing the speed of document review and reducing associated labor costs).

[220] FED. R. CIV. P. 26(b)(5)(A).

[221] *See* LEE & WILLGING, *supra* note 152 at 42-43 (reporting that the median portion of discovery costs was 1.6% of stakes for plaintiffs and 3.3% for defendants).

[222] *See supra* note 209 and accompanying text.

[223] *See supra* note 220.

empirical, particularly as TAR proliferates and the judicial response crystallizes. However, from the current vantage, and with appropriate humility about predicting technological change, TAR's proliferation is likely to progressively erode the cost asymmetries upon which the Court's *Twombly/Iqbal* doctrine is founded.

B. *Predictive Analytics and Forum Selection*

Forum selection offers a second concrete context in which to explore the intersection of legal tech and procedure. Indeed, legal tech firms are already marketing software that helps litigants choose the most advantageous forum in which to litigate their dispute. A leading example is Ravel Law, whose website features the following client testimonial: "With Ravel I can quickly perform a deep dive into how certain types of cases fare in a jurisdiction and the law that tends to control in a particular kind of case."[224] Surveys suggest substantial recent increases in use of data-based outcome-prediction tools among law firms.[225] In this Section, we seek to understand the possibilities, and also the significant limits, of outcome-prediction tools. In so doing, we offer a more skeptical take on legal tech's potential than in the e-discovery domain.[226]

Even so, focusing on forum selection offers an invaluable opportunity to probe legal tech's effect on the distribution of information within the system and explore how that might warrant a procedural response.

---

[224] *What People Are Saying*, RAVEL LAW, https://home.ravellaw.com [https://web.archive.org/web/20190207214629/https:/home.ravellaw.com/] (quoting Daniel Newman, Shareholder, Greenberg Traurig). Another example comes from a Shareholder at law firm Littler Mendelson:

> We are well on our way to being able to provide our clients with predictive analytics about case outcomes. If you have a case similar to one brought by a particular lawyer in a certain part of the country with the same judge, based on analytics we can predict the length of the case, the cost range and the possible outcome. That kind of information offers the power of prediction, and we serve our clients best when we can accurately predict outcomes and cost.

Lee Schreter, *In Their Words: Using Analytics and AI in Legal Practice*, GA. STATE UNIV.: GA. STATE NEWS HUB (Mar. 19, 2018), https://news.gsu.edu/2018/03/19/in-their-words-using-analytics-and-ai-in-legal-practice-2 [https://perma.cc/SE4Y-PBDS].

[225] *See* COALITION OF TECHNOLOGY RESOURCES FOR LAWYERS, DATA ANALYTICS IN CORPORATE LEGAL DEPARTMENTS: 2017-2018 TRENDS 6 (2018) (reporting a 43% increase in use by law firms of data analytics to perform outcome analysis and a 175% increase in anticipated spending on such tools).

[226] It is possible that predictive tools will be most useful not for forum-selection within the civil justice system but rather within the arbitration system, where litigants retain at least some agency in the selection of arbitrators. *See* Catherine A. Rogers, *Arbitrator Intelligence: From Intuition to Data in Arbitrator Appointments*, N.Y. DISP. RESOL. LAW., Spring 2018, at 41, 42 (noting use of tech tools to gain informational advantage in selecting arbitrators).

1. Forum-Shopping in Federal Courts and the Promise of Predictive
Analytics

A trio of features of the American litigation system has drawn entrepreneurial attention to forum selection as a legal tech target. First, the American system of federalism means that lawsuits can be heard in multiple fora, and a basic organizing principle is that plaintiffs in the U.S. civil justice system have the "venue privilege"—the right to choose the default place where a case is adjudicated.[227] Even so, defendants at both the federal and state level may move for statutory transfer to a new district or, in the federal system, seek dismissal from the federal system entirely using a common law forum non conveniens motion.[228] The result is that litigants on both sides of the "v" have a say in where a case is adjudicated.

Second, forum choice can have a significant impact on case outcomes, and so parties will have powerful strategic incentives to select or avoid particular fora by engaging in "forum-shopping." One set of incentives is the cost and convenience of litigating the suit. A party may hesitate to fight if forced to litigate in a faraway courthouse, or one that won't entitle the party to compulsory process for key witnesses.[229] Parties may also prefer a fast or slow resolution, and a forum that will facilitate it. Finally, a litigant's choice of forum can affect which law applies. In federal court cases involving state law claims, the court where a case is originated might affect choice of law because, under the Supreme Court's *Erie-Klaxon-Van Dusen* framework, courts typically apply the choice of law rule of the state where a civil action was removed or originally filed.[230]

Some of these differences are amenable to relatively low-cost analysis using conventional legal approaches—by reading cases and thinking like a lawyer. But forum choice can also matter for *how* law is applied. If advanced predictive analytics can be made to work in this arena—a very big "if" for reasons we discuss below—it would enjoy a decisive advantage. Regardless

---

[227] Atl. Marine Constr. Co. v. U.S. Dist. Ct. for the W. Dist. of Tex., 571 U.S. 49, 63 (2013).

[228] *See* 28 U.S.C. § 1404; Piper Aircraft Co. v. Reyno, 454 U.S. 235, 240-41 (1981).

[229] For classic examples, see *Carnival Cruise Lines, Inc. v. Shute*, 499 U.S. 585 (1991) and *J. McIntyre Machinery, Ltd. v. Nicastro*, 564 U.S. 873 (2011).

[230] Klaxon Co. v. Stentor Elec. Mfg. Co., 313 U.S. 487, 487 (1941) (holding that federal courts must follow the choice of law rules of the states in which they sit); *see also* Van Dusen v. Barrack, 376 U.S. 612, 633-634 (1964) (applying the choice of law rules of the filing state even if the case is transferred in the interests of convenience under § 1404(a)). *Klaxon* and *Van Dusen* further ensure that this is also true for removable cases initially filed in state court. However, there are two exceptions. If venue was improper in the original district, then following a transfer under 28 U.S.C. § 1406, the choice of law rules of the destination court's state will apply instead. In addition, the destination court's rules apply when an action is transferred to one designated in a forum selection clause. *See Atl. Marine Constr. Co.*, 571 U.S. at 66-67 (holding the parties to their forum-selection clause in the "interest of justice").

what law applies, judges in some jurisdictions might be more plaintiff-friendly than others in adjudicating motions to dismiss or for summary judgment, and the jury that awaits at trial if those motions are denied might be more generous.[231] The salience of these "discretionary" choices may grow over time amidst an increasingly politicized judiciary, as selected by an increasingly polarized political process,[232] and a jury pool shaped by Americans' growing tendency to sort along socioeconomic and ideological lines.[233]

A third feature of American litigation that makes predictive forum selection a potentially valuable growth area is that, by and large, American courts accept forum shopping as an intrinsic part of the system. An obvious exception, of course, is the *Erie* doctrine, which is explicitly structured around curtailing law-based incentives for forum-shopping as between federal and state courts.[234] But beyond *Erie*, and despite occasional judicial outbursts noting "the danger of forum shopping"[235] or declaring it "evil,"[236] the underlying doctrinal story, from the Supreme Court on down, is a far more accommodating one.[237] Part of this is a brute accommodation of the messiness of American federalism.[238] Part of it may be a determination that other procedural doctrines and statutes—among them personal jurisdiction, statutory limits on venue, and, as just noted, *Van Dusen*'s effort to ensure that

---

231 *See* Debra Lyn Bassett, *The Forum Game*, 84 N.C. L. REV. 333, 350 (2006) (providing examples of these more "subjective and personal factors"); *see also* Mary Garvey Algero, *In Defense of Forum Shopping: A Realistic Look at Selecting a Venue*, 78 NEB. L. REV. 79, 99 (1999) (suggesting that, even if the underlying law is the same, some courts are more likely to interpret and apply the law in a favorable way).

232 *See* NATHANIEL PERSILY, SOLUTIONS TO POLITICAL POLARIZATION IN AMERICA 5-6 (2015) (explaining the increasing polarization of contemporary American politics and exploring potential solutions); Richard L. Hasen, *Polarization and the Judiciary*, 22 ANN. REV. POL. SCI. 261, 262 (2019) (arguing that increased polarization affects both judicial selection and litigation outcomes).

233 *See, e.g.*, BILL BISHOP & ROBERT CUSHING, THE BIG SORT: WHY THE CLUSTERING OF LIKE-MINDED AMERICA IS TEARING US APART 1, 40 (2008) (noting the consequences of increasing ideological and socioeconomic segregation).

234 Erie R.R. Co. v. Tompkins, 304 U.S. 64, 78 (1938).

235 Agency Holding Corp. v. Malley-Duff & Assocs., 483 U.S. 143, 154 (1987) (discussing the danger of forum shopping as it pertains to RICO cases).

236 Wells v. Simonds Abrasive Co., 345 U.S. 514, 521 (1953) (Jackson, J., dissenting) (discussing how the application of the Full Faith and Credit Clause will not lead to the evil of forum shopping).

237 *See, e.g.*, Ferens v. John Deere Co., 494 U.S. 516, 527-28 (1990) (recognizing that plaintiffs engage in forum shopping); Van Dusen v. Barrack, 376 U.S. 612, 633-34 (1964) (same as to defendants); Keeton v. Hustler Mag., Inc., 465 U.S. 770, 779 (1984) (characterizing state-to-state forum shopping as a standard litigation strategy); Piper Aircraft Co. v. Reyno, 454 U.S. 235, 250 (1981) (same); *see also* Kermit Roosevelt III, *Choice of Law in Federal Courts: From Erie and Klaxon to CAFA and Shady Grove*, 106 NW. U. L. REV. 1, 32-33 (2012) (describing normative ambivalence about forum shopping). Interestingly, American ambivalence about forum shopping does not hold for global forum shopping. *See* Pamela K. Bookman, *The Unsung Virtues of Global Forum Shopping*, 92 NOTRE DAME L. REV. 579, 589, 592 (2016) (countering critics' view that international forum shopping is illegitimate).

238 *See* Friedrich K. Juenger, *Forum Shopping, Domestic and International*, 63 TUL. L. REV. 553, 569 (1989) (arguing that federalism permits or even invites forum shopping).

transfers on convenience grounds do not affect which law applies—place reasonable bounds around forum-shopping opportunities.[239] And part of it may be an artifact of a thoroughgoing adversarial system that sees litigation strategy, including forum-shopping, as synonymous with zealous representation and perhaps even an ethical duty.[240] Whatever the cause, even where litigants seek a venue transfer on convenience grounds, courts rarely scrutinize the deeper strategic purpose that that request often reflects.

2. Will Predictive Forum Selection "Work"?

While some believe predictive analytics methods hold great promise for litigants seeking to maneuver their dispute into an advantageous forum, serious questions, unrelated to the NLP challenges discussed previously, remain as to legal tech's ability to deliver on any such promise. These concerns may or may not be insuperable. At a bare minimum, they indicate that strong headwinds must shape thinking about any procedural response.

Start with a concrete example: whether a defendant in an already-filed case should move to transfer to another forum, given that the defendant plans to move to dismiss for failure to state a claim. In this setting, a defendant might want to know what share of all Rule 12(b)(6) motions has been granted in each district. A more refined approach would filter cases by additional available details, such as the PACER-reported nature of suit code, the number of parties on each side, their corporate status, the number of claims filed, the presence of state or federal law questions, and the court and assigned district court judge.[241] Machine learning methods can determine which, if any, of these variables importantly predict the result of Rule 12(b)(6) motions among the universe of already-litigated cases. If the district court is one of the important predictors, then a transfer to a more favorable district might be a good bet.

This simple example surfaces a key criterion for thinking about what it means for predictive analytics to "work": Available data must be useful in predicting the ways important case outcomes would vary across districts. Call this the *APU criterion*—the requirement that A̲vailable data is P̲redictively U̲seful.

---

[239] *See* Bassett, *supra* note 231, at 369 (contrasting procedural restrictions on forum-shopping with a more "laissez-faire approach"). For key personal jurisdiction cases, see *Goodyear Dunlop Tires Operations, S. A. v. Brown*, 564 U.S. 915 (2011); *Daimler AG v. Bauman*, 571 U.S. 117 (2014); *Bristol-Myers Squibb Co. v. Superior Ct.*, 137 S. Ct. 1773 (2017).

[240] Bassett, *supra* note 231, at 344 ("Indeed, the failure to forum shop would, in most instances, constitute malpractice.").

[241] Note, however, that the Nature of Suit (NOS) codes selected by each plaintiff in completing a civil cover sheet have been found to be problematic guides. *See* Christina L. Boyd & David A. Hoffman, *The Use and Reliability of Federal Nature of Suit Codes*, 2017 MICH. ST. L. REV. 997, 1006-07 (observing that the attorneys who select NOS codes have no training or standardized guidance and little incentive for selecting proper codes). This means that a predictive analyst would likely need to oversample cases to ensure she is not filtering out cases with relevant claims.

Predictive analytics applied to forum selection could fail the APU criterion in any of three ways. The first is insufficient data of the right type. While many of the variables relevant to predicting case outcomes are available in the docket reports that reside on the federal courts' PACER e-filing system, capturing the key case features with respect to, say, the plausibility standard applied to motions to dismiss under *Twombly/Iqbal* might vary in ways that require a wider catalog of case materials—for instance, complaints, memoranda of law supporting a motion to dismiss, or other documents. One problem is that PACER's search interface, which has all the sophistication and user-friendliness of its mid-1990s design, makes it almost useless for data filtering. The more significant issue is that, even if efficient filtering were possible, PACER fees, assessed on a document-by-document basis, would mount quickly in any effort to generate enough observations to support viable machine learning methods.[242] Federal-level court data, as one pair of scholars memorably put it, sits behind "a wall of cash and kludge."[243]

Could some alternative mechanism arise, duplicating PACER's massive holdings and allowing smart sharing of case documents? Westlaw, Lexis, and Bloomberg already download PACER docket reports and large numbers of underlying case documents. And large law firms surely possess expansive document collections they have filed and downloaded in their own work. There is also the insurgent RECAP archive, which makes freely available any document the archive's users have paid PACER to download.[244]

But there is little way to know how well some of these document collections represent the full population of cases.[245] And foreboding economics give good reason to doubt that any such collection will become comprehensive. One estimate found several years ago that the cost of downloading all of PACER

---

242 If, based on a power analysis, a litigant needs to download information on 20,000 cases to make useful predictions, and each case featured an average of three documents at an average cost of $1.50, total costs—for one case—could exceed $90,000, not including the costs of employing data scientists.

243 Charlotte S. Alexander & Mohammad Javad Feizollahi, *On Dragons, Caves, Teeth, and Claws: Legal Analytics and the Problem of Court Data Access*, *in* COMPUTATIONAL LEGAL STUDIES: THE PROMISE AND CHALLENGE OF DATA-DRIVEN LEGAL RESEARCH 97 (Ryan Whalen ed., 2019).

244 *Advanced RECAP Search*, COURT LISTENER https://www.courtlistener.com/recap [https://perma.cc/XSG9-5YLZ]. RECAP, for those who missed it, is "PACER" spelled backward.

245 Even Westlaw and Lexis have gaps. *See* McAlister, *supra* note 37 (describing the incompleteness of legal databases); Elizabeth Y. McCuskey, *Submerged Precedent*, 16 NEV. L.J. 515, 517 (2016) (finding that Westlaw and Lexis lacked roughly thirty percent of reasoned decisions by judges in two federal district courts). Harvard Law School's noble Caselaw Access Project claims to eliminate these gaps, compiling and making freely available every published decision in the history of American law. *See* CASELAW ACCESS PROJECT, https://case.law [https://perma.cc/5BV6-XHZA]. But even this is insufficient for many legal analytics applications, since published decisions are only the tip of the iceberg of litigation. *See* Engstrom, *supra* note 214, at 1208-09 (2013) (studying published and unpublished decisions to draw conclusions on civil processes).

would have been as much as $1 billion.[246] And that estimate doesn't include the cost of updating the data pool with the tens of millions of additional documents filed in the federal courts each year going forward.[247] In economics terms, PACER enjoys a natural monopoly and seems uninclined to relinquish that position, making it hard for new entrants to gain a foothold.[248] Further, the magnitude of data costs suggests that, even if an entity were willing to invest, the tools that ultimately made it to market would not likely be made widely available. Indeed, the more likely outcome is that litigation's "have nots" will be priced out, particularly since the value of the tool to litigation's "haves," and thus the price they are willing to pay for it, will derive at least in part from exclusive access to its outputs. That leaves the possibility of public sector-driven

---

[246] Michael Lissner, *The Cost of PACER Data? Around One Billion Dollars*, FREE L. PROJECT (Oct. 10, 2016), https://free.law/2016/10/10/the-cost-of-pacer-data-around-one-billion-dollars [https://perma.cc/C9HM-C5AG].

[247] John Brinkema & J. Michael Greenwood, *E-Filing Case Management Services in the US Federal Courts: The Next Generation: A Case Study*, INT'L J. CT. ADMIN., Summer 2015, at 3, 3. The problem is ever greater at the state level, where filings are even higher. In Florida—to choose at random a relatively large state—litigants filed roughly 25.5 million documents, totaling about 119 million pages, between mid-2018 and mid-2019 alone. FL. CTS. E-FILING AUTH. 2018-2019 ANNUAL STATISTICS (2019), https://www.myflcourtaccess.com/authority/e-Filing_Authority/Resources/2019-2020_Board _Meetings/Annual_Report_2019/Florida_Courts_E-Filing_Authority_Annual_Report-2018-19.pdf [https://perma.cc/6Y3R-7CKA].

[248] Natural monopolies are characterized by declining average costs as output increases, meaning high fixed and low marginal costs. PACER is an instance because the fixed costs of collecting and indexing the data dwarf the marginal costs of searching and sharing it over the internet. *See* Prateek Agarawal, *Natural Monopoly*, INTELLIGENT ECONOMIST, https://www.intelligenteconomist.com/natural-monopoly [https://perma.cc/7DNG-Z8AE] (Feb. 7, 2020). To concretely play out PACER's potential invulnerability to competition, suppose that PACER generates about $150 million in annual fees. This is an upper bound for recent years. *Compare* U.S. CTS., FY 2020 CONGRESSIONAL BUDGET REQUEST: JUDICIARY INFORMATION TECHNOLOGY FUND 11.2 tbl.11.1 (2019), https://www.uscourts.gov/about-federal-courts/ governance-judicial-conference/congressional-budget-request [https://perma.cc/GQJ7-X9BN] (reporting fiscal year 2019 and 2020 estimates of $147.7 million in "Estimated Receipts and Prior Year Recoveries" for the "EPA Program" (the "Electronic Public Access" program), which encompasses PACER) *with* U.S. CTS., FY 2022 CONGRESSIONAL BUDGET REQUEST: JUDICIARY INFORMATION TECHNOLOGY FUND 11.2 tbl.11.1 (2021), https://www.uscourts.gov/about-federal-courts/governance-judicial-conference/congressional -budget-request [https://perma.cc/7Z6Q-5FNW] (reporting $156.8 million in the same category for fiscal year 2020 and estimating $144.5 million for fiscal years 2021 and 2022). If a competitive entrant could buy a stream of revenue of $150 million for "only" $1 billion, it would be profitable to do so as long as the entrant's next-best investment yielded returns below 15%, or $150 million. This, of course, ignores operating costs, but there is reason to think such costs, including data warehousing and bandwidth, would be low. *See* Lissner, *supra* note 246, at n.3 (estimating a cost of $128,000 annually for data warehousing); *Comparing Bandwidth Costs of Amazon, Google and Microsoft Cloud Computing*, ARADOR (May 3, 2017), https://arador.com/ridiculous-bandwidth-costs-amazon-google-microsoft [https://perma.cc/R9ZB-YT3Z] (suggesting bandwidth costs of roughly $100 per TB per month). But, as noted previously, the entity would also have to download from PACER tens of millions of additional documents each year. And there is no guarantee—indeed, plenty of reason to doubt—that PACER would continue to operate if a competitor gobbled up its revenue stream. With no PACER, there would be no bulk source of federal court data. And this ignores the possibility that new entrants would appear; once they have sunk the entry costs, price competition could render *all* providers unprofitable.

reform, perhaps as a result of litigation challenging PACER's policies or because the Judicial Conference or Congress steps in.[249] Short of this, however, data limitations may well place a ceiling on predictive forum selection.

A second way that predictive analytics applied to forum selection might fail the APU criterion derives from a particular kind of endogeneity. Using predictive analytics to drive forum selection decisions might well cause changes in litigant behavior that erode any initial accuracy or usefulness. This is an instance of the "Lucas critique"—named for economist and Nobel laureate Robert Lucas.[250] Put in simplest terms, systematic patterns in litigation outcomes reflect endogenous strategic behavior by litigants. Patterns revealed by deployment of predictive analytics methods can be expected to induce behavior changes *as a result of the use of predictive analytics methods themselves*. This, in turn, might destroy the future accuracy of the very prediction methods that drove the change in behavior.

Here's an example of how the Lucas critique problem might operate. One core feature of concern to litigants is the amount of time a case takes to wend its way to the finish line. If predictive analytics indicate that a particular forum is better for parties with the ability to select it, then parties will flock to this "magnet" forum, clogging up its docket, thereby slowing down *all* litigation there. The opposite will happen in "source" forums. In principle, Congress could respond to such a result by increasing the number of judgeships in magnet forums. But that would take time, and it presumes that Congress would act for efficiency's sake, which may not be realistic given the current political climate around judgeships.[251] This example of source-magnet dynamics shows how behavioral changes could endogenously reduce the value of the information gained.

---

[249] *See* Re & Solow-Niederman, *supra* note 3, at 285 (proposing a "public option" legal tech tool and accompanying data). On litigation disputing PACER fees, see *Greenspan v. Admin. Off. Of the U.S. Cts.*, No. 14-2396, 2014 WL 6847460, at *3 (N.D. Cal. 2014); *Fisher v. Duff*, No. 15-5944, 2016 WL 3280429, at *1 (W.D. Wash. 2016); *Fisher v. United States*, 128 Fed. Cl. 780, 783 (2016); *Nat'l Veterans Legal Servs. Program v. United States*, 291 F. Supp. 3d 123, 126 (D.D.C. 2018). In the fourth of these, the Federal Circuit recently upheld the lower court's judgment, finding that the courts were charging excessive fees in violation of the PACER authorizing legislation and the E-Government Act of 2002. *See Nat'l Veterans Legal Servs. Program v. United States*, 968 F.3d 1340, 1350 (Fed. Cir. 2020) (holding that fees must be limited to the expenses "incurred in services providing public access"). Calls for more open court data are not new. *See* Alexander & Feizollahi, *supra* note 243 (recounting history of scholarly and other activism). *See also* Adam R. Pah, David L. Schwartz, Sarath Sanga, Zachary D. Clopton, Peter DiCola, Rachel Davis Mersey, Charlotte S. Alexander, Kristian J. Hammond & Luís A. Nunes Amaral, *How to Build a More Open Justice System*, 369 SCIENCE 134, 136 (2020) ("[O]rganizations outside government should directly purchase and publicize court records.").

[250] *See* Robert E. Lucas, Jr., *Econometric Policy Evaluation: A Critique*, 1 CARNEGIE-ROCHESTER CONF. SERIES ON PUB. POL'Y 19, 41-42 (1976) ("[U]nannounced sequences of policy decisions . . . appear[] to be beyond the capability not only of the current-generation models, but of conceivable future models as well.").

[251] *See, e.g.*, Hasen, *supra* note 232 (discussing the increased polarization of the judiciary).

A third issue, which also reflects endogeneity, operates prior to the Lucas critique problem. Ever since Priest and Klein's seminal article, it has been a commonplace assumption that the set of cases that make it to judgment is systematically selected.[252] Some cases settle before trial, and it is unlikely to be random which ones do.[253] A reasonable conclusion to draw is that cases for which we observe litigation outcomes differ from cases that settle before those outcomes would be observed, as well as from cases in which those outcomes never would be observed.

To make this more concrete, suppose parties have access to two forum options, A and B, of roughly equal size. All cases are diversity cases, and each involves one Forum A party and one Forum B party. Without predictive analytics, suppose it is essentially random where cases are heard in the sense that the plaintiff just files where she lives, and defendants move to transfer in some but not all cases. Now assume that some defendants gain access to predictive analytics. They find that motions to dismiss are granted 60% of the time in tort cases heard in Forum A but only 20% of the time in Forum B. With predictive analytics, (i) all tort defendants will decline to seek transfer out of Forum A and (ii) all tort defendants will seek transfer out of Forum B. In short, the advent of predictive analytics causes many more tort cases to be heard in Forum A, and many fewer in Forum B. But 12(b)(6) grant rates will remain three times greater in Forum A only if the cases newly litigated in Form A are similar to those previously heard there. If, however, pre-analytics Forum B tort cases were stronger or better pleaded than in Forum A, then the grant rate will not be 60%. Because the parties' strategic choices will shape

---

[252] *See* George Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 133 J. LEGAL STUD. 1, 4 (1984) (presenting a model in which the determinants of litigation are purely economic, such as direct costs of litigation and rational estimates of the outcome). For more recent revisions, see Daniel M. Klerman & Yoon-Ho Alex Lee, *Inferences from Litigated Cases*, 43 J. LEGAL STUD. 209, 211-12 (2014) (discussing the impact of pro-plaintiff and pro-defendant legal standards on litigation); Jonah B. Gelbach, *The Reduced Form of Litigation Models and the Plaintiff's Win Rate*, 61 J.L. & ECON. 125, 150 (2018) [hereinafter Gelbach, *Reduced Form*] (demonstrating the flexibility of the Priest-Klein framework); Eric Helland, Daniel Kerman & Yoon-Ho Alex Lee, *Maybe there Is No Bias in the Selection of Disputes for Litigation*, 174 J. INSTITUTIONAL & THEORETICAL ECON. 143, 143-44 (2018) (analyzing data from contingent-fee lawyers in New York); Jonah B. Gelbach, Commentary, *Maybe There is No Bias in the Selection of Disputes for Litigation*, 174 J. INSTITUTIONAL & THEORETICAL ECON. 171, 171 (2018) (arguing that the data "indicate considerably less similarity across adjudicated and settled cases" than previously believed).

[253] This same logic can be applied in other areas of pre-trial litigation, including, e.g., 12(b)(6) and summary judgment. With respect to the 12(b)(6) stage, see, e.g., Gelbach, *supra* note 216, and Issacharoff & Miller, *supra* note 216; with respect to summary judgment, see Jonah B. Gelbach, *Rethinking Summary Judgment Empirics: The Life of the Parties*, 162 U. PA. L. REV. 1663 (2014). Other examples include removal, see, e.g., Kevin M. Clermont & Theodore Eisenberg, *Do Case Outcomes Really Reveal Anything About the Legal System? Win Rates and Removal Jurisdiction*, 83 CORNELL L. REV. 581 (1998), and patent litigation, see, e.g., Jonathan Masur, *Patent Inflation*, 121 YALE L.J. 470, 514 (2011).

observed outcomes, it will be difficult to lay down clear and verifiable conditions under which win rates are predictable.

Thus, whereas the Lucas critique suggests that initially valuable prediction methods will induce behavioral changes that destroy predictive usefulness, the selection problem Priest and Klein describe might render the initial predictions too inaccurate to be useful in the first place. Both sources of endogeneity support healthy skepticism of the ability of predictive analytics methods to "work" in guiding forum selection choices.

Perhaps all is not lost for legal tech entrepreneurs, because analysts could try modeling endogenous behavior directly. Successful estimation of what economists call "structural models" of behavior would allow predictions that are robust to both the Lucas and Priest-Klein forms of endogeneity described above.[254] But such estimation typically must rely critically on contestable behavioral and statistical assumptions. However, those are rarely the focus of predictive analytics methods, which are usually regarded as a black-box-ish alternative to structural modeling.[255] Still, it is at least possible that, as predictive analytics methods proliferate and become pervasive, the system will reach a more-or-less stable equilibrium.[256] If so, and if enough people behave in ways in line with what predictive analytics indicate—possibly *because* of those predictions—then the predictions may turn out to be right in equilibrium. So long as no large shocks hit the system, predictions would then be useful. But overall, the endogeneity of litigant behavior poses yet another

---

[254] Structural models "focus on distinguishing clearly between the objective[s] . . . of the economic agents and their opportunit[ies] . . . as defined by the economic environment." Hamish Low and Costas Meghir, *The Use of Structural Models in Econometrics*, 31 J. ECON. PERSPS. 33, 33 (2017). By using clearly stated assumptions, researchers who deploy and estimate structural models are able to "identify[] mechanisms that determine outcomes," which allows them to "analyze counterfactual policies, quantifying impacts on specific outcomes as well as effects in the short and longer run." *Id.* at 33-34.

[255] A researcher focused on predicting the value of a variable implicitly assumes that the existing pattern of behavioral relationships will persist during the period when the prediction will be used. This means it isn't important *why* the prediction is or isn't accurate—all that matters is the accuracy itself. *See, e.g.*, Kleinberg et al., *supra* note 31, at 493 ("Machine learning techniques . . . provide a disciplined way to predict [variables]."). By contrast, structural modelling's focus is on clarifying and quantifying causal mechanisms to enable useful predictions under *different* conditions. *See supra* note 254. This is not to say there is no overlap between methods used for prediction and those used for causal inferences; for an example involving securities litigation, see Andrew C. Baker & Jonah B. Gelbach, *Machine Learning and Predicted Returns for Event Studies in Securities Litigation*, 5 J.L. FIN. & ACCT. 231, 233 (2020) (explaining that "event studies, as used in securities litigation, can be viewed as out-of-sample prediction problems").

[256] In mathematical terms, the iteration of predictions and forum choices might function together as a "contraction mapping," causing the system to converge to a stable equilibrium. *See, e.g.*, Robert M. Brooks & Klaus Schmitt, *The Contraction Mapping Principle and Some Applications*, ELEC. J. DIFFERENTIAL EQUATIONS, Monograph 09, 2009, at 2 (explaining the contraction mapping principle in a way that involves conditions under which an end result "may be obtained as the limit of an iteration scheme" from "an arbitrary starting point").

significant technical barrier to the world of robojudges and robolawyers imagined in much of the existing legal tech literature.

3. The Future of Forum Selection and Civil Procedure

The above discussion provides grounds for skepticism that predictive forum selection will gobble up the litigation world. But it is at least possible that it will "work" well enough to support a robust market for its use. What are the implications if legal tech tools applied to forum selection turn out to be predictively useful, and reliably so? We address two possibilities.

First, the emergence of robust outcome prediction tools and a consequent rise in digitized forum selection game-playing could revive concerns among judges— and, eventually, rulemakers and policymakers—about "manipulable justice" and "unprincipled gamesmanship" that have otherwise largely fallen away within the American system.[257] If forum-shopping lost legitimacy as an intrinsic part of the litigation landscape, a reformist impulse might break through.

The opening of policy windows is rarely a given. Of particular importance will be the cogency of the empirical showings that can be made—about, say, the volume of satellite litigation or the degree to which litigation's "haves" are systematically gaining advantage over its "have nots."[258] So would evidence of a sharpening of what some see as a worrying practice of courts openly competing for business by offering procedural or other carrots—a phenomenon that likely fueled recent patent venue reforms.[259] A shift in forum shopping's valence might also gain momentum from interventions elsewhere, such as France, where warnings about the "Ravelization of law" accompanied recent legislation banning judicial analytics and, indeed, imposing criminal penalties for their use.[260] To be sure, the French reaction can be chalked up to its status

---

[257] Bassett, *supra* note 231, at 388; Bookman, *supra* note 237 at 579.

[258] *See* Ori Aronson, *Forum by Coin Flip: A Random Allocation Model for Jurisdictional Overlap*, 45 SETON HALL L. REV. 63, 75-76 (2015) (discussing the disparities between parties with more resources to gather and use information in forum-shopping); Michael H. Gottesman, *Draining the Dismal Swamp: The Case for Federal Choice of Law Statutes*, 80 GEO. L.J. 1, 14 (1991) ("As corporations often have greater financial resources than their 'victim' adversaries, they are likely to be more efficient at utilizing that opportunity and to win a disproportionate number of the races [to the courthouse of their choice]."); Kevin M. Clermont & Theodore Eisenberg, *Exorcising the Evil of Forum-Shopping*, 80 CORNELL L. REV. 1507, 1511-12 (1995) (describing the disparities in outcome between cases that transfer venue and cases that stay put).

[259] *See* J. Jonas Anderson, *Court Competition for Patent Cases*, 163 U. PA. L. REV. 631, 634 (2015) (describing how courts compete for patent litigants with procedural carrots); Daniel Klerman & Greg Reilly, *Forum Selling*, 89 S. CAL. L. REV. 241, 244 (2016) (discussing how courts cater to plaintiffs to attract litigants); Scott E. Atkinson, Alan C. Marco & John L. Turner, *The Economics of a Centralized Judiciary: Uniformity, Forum Shopping, and the Federal Circuit*, 52 J.L. & ECON. 411, 411-12 (2009) (studying the impact of the creation of the Federal Circuit on uniformity of outcome in patent cases).

[260] *See* Jason Tashea, *France Bans Publishing of Judicial Analytics and Prompts Criminal Penalty*, A.B.A. J. (June 7, 2019, 1:51 PM), https://www.abajournal.com/news/article/france-bans-and-creates-

as a civil code country committed to ex ante codification of law.[261] In common law systems founded on judge-made decisional law, the threat that predictive analytics will yield a legal realist unmasking of law's politics and indeterminacies is less acute, if only because so many observers already accept the legal realist view.[262] That said, the France example may also reflect a growing and more universal distrust, particularly in democratic systems, of use of algorithmic decisionmaking throughout society.[263]

Forum shopping's demotion will be important because it is likely to be a precondition of a second possible consequence: with forum-shopping's valence flipped in the judicial or legislative mind, either type of actor might take action to reform the system. Congress could, à la France, prohibit use of predictive analytics for forum selection purposes. It could also revise the federal venue statute to narrow or outright eliminate transfers on pure convenience grounds. Either of these approaches, however, brings obvious challenges. The former would be hard, if not impossible, to police. The latter would be hard to maneuver through the current Congress, or any Congress, since any constriction of venue transfer would systematically disadvantage defendants, often corporate ones.[264]

---

criminal-penalty-for-judicial-analytics [https://perma.cc/KJG5-A7KG] (reporting that France imposed a criminal penalty of up to five years in prison for publication of judicial analytics). The legislative episode was somewhat more complex than the American legal press suggested, as France first passed a pioneering law establishing fully open court data and then, when judges balked, enacted the prohibition. Alschner, *supra* note 78 (manuscript at 12).

261 J. Lyn Entrikin, *The Death of Common Law*, 42 HARV. J.L. & PUB. POL'Y 351, 433-43 (2019) (providing a comprehensive overview of the distinction between common law and civil law legal systems).

262 *Id.* at 442-43 (noting differences in "constraints on judicial discretion" and "legal development" across the two systems, with common law systems following an "incremental" and "case by case" approach that embraces judicial policymaking, and civil law systems adhering to the idea that judicial decisions are "persuasive but never binding" and a "'gloss' [on] codified law," with the code providing the "authentic statement of fundamental principle" (quoting F.H. Lawson, *A Common Law Lawyer Looks at Codification*, 2 INTER-AM. L. REV. 1, 5 (1960))); *see also* Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, *in* A MATTER OF INTERPRETATION 10 (Amy Gutmann ed., new ed. 2018) ("It is only in this [20th] century, with the rise of legal realism, that we came to acknowledge that judges in fact 'make' the common law . . . .").

263 *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2016) (discussing the problems of privacy and transparency in an increasingly tech-driven society).

264 *See supra* note 258 and accompanying text. Equally unlikely, and more the stuff of academic inquiry, is a statutorily prescribed *randomized* allocation system. *See* Aronson, *supra* note 258, at 66 (proposing a lottery system allocate cases among jurisdictions). Still another possibility is that Congress could enact a federal choice-of-law statute to reduce the advantage plaintiffs can gain by filing in a forum to get their preferred choice of law under *Van Dusen*. Note, however, some problems. Many have questioned whether choice of law is amenable to statutory codification at all, and most proposals coming out of the last round of anxiety about forum-shopping, in the 1990s, advanced a thicket of competing canons tailored to specific subjects or types of collisions between legal rules. *See, e.g.*, Larry Kramer, *Rethinking Choice of Law*, 90 COLUM. L. REV. 277, 322-40 (1990) (recommending

Instead, the most likely procedural response to an escalation in effective predictive forum selection practices will come from judges, not legislators. And given this, the most likely intervention over the near- to medium-term will not take the form of legislative amendments to the venue rules but rather a regime of *disclosure*, via judicial demands for litigants' machine outputs.

What, precisely, would this look like? Consider several options. Judges facing a transfer motion could require the parties to disclose the fact of their use of predictive analytics. More aggressively, parties could be required to disclose to all sides, including the judge, their models' predictions for each forum they considered. Most aggressive of all would be a requirement that a party who uses predictive analytics give direct access to the programs and/or code used to generate predictions. Disclosure could, in turn, lead to the crafting of new rules, whether by judges or via the rulemaking process, distinguishing types of reasons surfaced via predictive analytics. Some predictions could be treated as affirmative reasons for transfer, akin to reduced litigation costs. That would make sense in the case of timing-related predictions. After all, it may be less costly to litigate in a district where, all else equal, the case moves more quickly. By contrast, predictions related to who will win dispositive motions relate to a zero-sum variable, so the associated "convenience" for one party is "inconvenience" for the other. The case for transfer in such cases turns importantly on distributional considerations— which party do we want to favor?—rather than on efficiency-based arguments.

In the current political climate, a judge-made disclosure regime is more realistic than Congress tweaking the venue statute, but it is also, in a system founded upon adversarialism, more bracing. Compelled disclosure of machine outputs or source code would implicate the anti-free-riding justification for work product protection.[265] For now, however, it is worth noting that there remain many further questions about whether and when a disclosure regime would make sense as a policy matter.

For instance, a threshold question—and one we also return to later—is whether judge-litigant or litigant-litigant information asymmetry is the critical challenge. In cases with sophisticated, well-financed parties on all sides, perhaps adversarialism will take care of judge-litigant information

---

a series of different canons for different types of cases); *see also* LEA BRILMAYER, CONFLICT OF LAWS: FOUNDATIONS AND FUTURE DIRECTIONS 161-67, 185-89 (1991) (evaluating reciprocity requirements, uniform legislation and the restatements published by the American Law Institute as potential "choice of law solutions"); Gottesman, *supra* note 258, at 1 (suggesting legislative enactment of choice of law rules for multistate litigation). The bigger problem is that a unified choice-of-law regime might not accomplish much if, as noted previously, predictive forum selection proves most useful in exploiting the ideology-inflected decisions of an increasingly politicized judiciary and demographically sorted jury pools. *See supra* notes 231–233 and accompanying text. After all, these choices operate within law's interstices; they do not depend on choices among legal rules.

265   We systematically address issues related to the work product doctrine momentarily, in Section II.C.

asymmetry. If both sides have access to the same quality predictions, then at least one of them will have the incentive to inform the court that predictive analytics likely motivates the quest for a change of venue. Thus, litigant-litigant information asymmetry—itself likely to result from *litigant-litigant resource* asymmetry—is ultimately the source of *judge-litigant information* asymmetry. This interesting result indicates that predictive forum selection, if it comes to be disfavored, likely requires active policing by judges only in the presence of significant litigant resource disparities—that is, only when litigation's "haves" and "have nots" face off.

Another key policy question is whether compelled disclosure would chill use of predictive analytics for forum selection and whether we should care. For example, disclosure might induce some defendants—particularly those with pre-existing knowledge about a "magnet" district's desirability—not to use predictive analytics at all. Our hunch is that this should not matter: It is in middle-ground cases where forum choices are less obvious that defendants would use analytics even when forced to disclose, and these are, by construction, the cases where analytics are likely most valuable to defendants. Even so, more thinking will clearly be required to work through the costs (e.g., distributive concerns across litigation's "haves" and "have nots") as against its benefits (e.g., earlier and potentially socially efficient settlements[266]). That cost-benefit comparison, and the many other research questions flagged above, will provide fruitful avenues for further inquiry as predictive forum selection tools improve and their procedural regulation comes into clearer focus.

C. *From Borrowed Wits to Borrowed Bits: Legal Tech and the Work Product Doctrine*

This section turns to an issue that has lurked in the background of the analysis to this point: the treatment of legal tech tools, including TAR and predictive forum selection tools but also tools that perform advanced legal analytics, under the work product doctrine.

---

[266] Suppose predictive analytics tells its users where each party is most likely to win. One view might be that it involves nothing but a redistribution from plaintiffs to defendants. But that's too facile. After transfer, both parties might become certain the defendant would win, so the case is likely to settle, reducing both private and public litigation costs. Presumably the settlement would be on poor terms for the plaintiff, so now we have a tradeoff between normative considerations related to the plaintiff's loss of bargaining power and litigation-cost considerations related to early settlement. Such effects on settlement behavior greatly complicate any attempt to predict the net benefits of changes in litigation policy. *See* Jonah B. Gelbach, *Can the Dark Arts of the Dismal Science Shed Light on the Empirical Reality of Civil Procedure?*, 2 STAN. J. COMPLEX LITIG. 223, 292 (2014) (discussing the problems with employing empirical research to evaluate changes to procedural questions); Jonah B. Gelbach, *Rethinking Summary Judgment Empirics: The Life of the Parties*, 162 U. PA. L. REV. 1663, 1668-69 (2014) (describing empirical evidence consistent with the claim that parties change their settlement demands in response to changes in the context of the trial).

1. Information and Adversarialism: Reframing Legal Tech's Distributive Costs

Legal tech, we noted way back in Part I, is likely a double-edged sword as a distributive matter. On one hand, it can narrow adversarial inequities by providing a force multiplier to under-resourced counsel and by making legal redress available to categories of claimants who are not served, or poorly served, within the current system.[267] On the other hand, legal tech can deepen distributive divides because, among other things, the "haves" may be better positioned to capture legal tech's efficiencies and then use them to deploy more law, and deploy law more effectively, against the "have nots" rather than the other way around.[268]

These are important and interesting possibilities that will surely repay further research as legal tech proliferates. But command of the full landscape of legal tech and some of its technical possibilities and limits also permits a more focused and concrete set of claims about legal tech's likely distributive consequences. In particular, virtually every tool in the legal tech toolkit aspires to confer on users better information than their adversaries—about the best forum in which to litigate, the most damaging documents in a vast production, the likelihood of winning before *this* judge or jury, and the best arguments to lay before either actor to get there. It follows that, as various tools within the legal tech toolkit improve, and if only the "haves" can access the best of them, one could expect a widening of information asymmetries—whether in particular litigation areas, or even across the system as a whole—that will exacerbate, rather than mitigate, distributive concerns and permit some groups to systematically win out over others.

Consider two concrete examples. First, a key question in emerging e-discovery debates is whether TAR increases or decreases gaming and abuse.[269] TAR's champions hold that it can replace human subjectivity and bias with the

---

267 See *supra* notes 101–105, 131–135 and accompanying text.

268 *Id.*

269 For an overview of longstanding debate about discovery abuse's prevalence, see Linda S. Mullenix, *Discovery in Disarray: The Pervasive Myth of Pervasive Discovery Abuse and the Consequences for Unfounded Rulemaking*, 46 STAN. L. REV. 1393 (1994). Empirical studies have also surveyed the debate. *See* AM. COLL. OF TRIAL LAWS. & INST. FOR THE ADVANCEMENT OF THE AM. LEGAL SYS., INTERIM REPORT ON THE JOINT PROJECT OF THE AMERICAN COLLEGE OF TRIAL LAWYERS TASK FORCE ON DISCOVERY AND INSTITUTE FOR THE ADVANCEMENT OF THE AMERICAN LEGAL SYSTEM 4 (2008), https://iaals.du.edu/sites/default/files/documents/publications/interim_report_final_for_web.pdf [https://perma.cc/W7BH-A93D] (reporting that 45% of lawyers surveyed believed discovery abuse occurred in "almost every case"); INST. FOR THE ADVANCEMENT OF THE AM. LEGAL SYS., CIVIL CASE PROCESSING IN THE FEDERAL DISTRICT COURTS: A 21ST CENTURY ANALYSIS 46 (2009) (finding discovery sanctions are filed in about three percent of cases and imposed in twenty-six percent of the cases in which they were filed); LEE & WILLING, *supra* note 152, at 14 (same).

"mechanical objectivity" of a machine.[270] TAR also leaves a decisionmaking trail with methods and models and, taking a page from the wider algorithmic accountability literature, means that litigants must "show their work" in ways that can increase transparency relative to analog approaches.[271] However, TAR may also increase gaming opportunities.[272] Indeed, better-heeled parties can construct seed sets and make modeling choices they know will yield fewer relevant documents and exclude especially harmful ones. Many of these artifices, embedded deep in code, will likely go unnoticed and unchallenged, particularly where less sophisticated parties sit on the other side. Even where sophisticated litigants negotiate ex ante a protocol governing seed set construction, statistical methods, and back-end evaluation and validation techniques,[273] the opacity of algorithmic outputs and the hands-on nature of training and tuning machine learning models can deprive TAR systems of basic "contestability."[274] Far from bringing transparency and "mechanical objectivity," automated discovery might breed *more* abuse, and prove less amenable to oversight, than an analog system built upon "eyes-on" review.

A second example focuses on a type of legal tech tool that has not yet occupied much of the discussion to this point: legal analytics tools that help a litigant predict a case's resolution, not for forum-shopping purposes, but to inform a party's settlement calculus and litigation strategy once a case sits before a particular judge. These tools, we noted previously, are currently most advanced in technical, self-contained areas of law like tax and employment, but they are likely to branch out, particularly as entities with privileged data access—large, repeat institutional players like Walmart, or law firms that specialize in particular litigation—use their privileged access to data to develop potent analytics tools in less siloed areas.[275] The distributive concern

---

[270] Remus, *supra* note 34, at 1701.

[271] The analogy here is the "algorithmic accountability" literature's argument that algorithmic tools may render decisions—of employers, manufacturers, governments—*more* transparent than their analog versions. *See, e.g.*, David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REGUL. 800, 804 (2020) (arguing that "algorithmic governance" could bring more, not less, transparency and accountability to the administrative state); Cass R. Sunstein, *Algorithms, Correcting Biases*, 86 SOC. RSCH. 499, 510 (2019) (arguing that algorithms force "unprecedented transparency" by requiring people to make deliberate choices about policy tradeoffs).

[272] E-discovery is already an area that many see as rife with abuse. Dan H. Willoughby, *Sanctions for E-Discovery Violations: By the Numbers*, 60 DUKE L.J. 789, 790-91 (2010) (finding a sharp uptick in motions seeking and awards of discovery sanctions).

[273] Some say reliability and quality control measures (e.g., confidence intervals, prediction score thresholds) are either slippery or suffer from a lack of ground truth (e.g., a baseline relevance rate), limiting their ability to discipline the responding party. *See* Endo, *supra* note 34, at 854.

[274] *Id.* at 863 (detailing how the "black-box" quality of predictive coding makes it harder for less sophisticated litigants to challenge the predictive coding process); *see also* Hildebrandt, *supra* note 128, at 29-30 (arguing for the necessity of contestability conditions); Kluttz & Mulligan, *supra* note 105, at 884-88(calling for new approaches to "validation and testing" focused on "contestability" of design).

[275] *See supra* notes 140–144 and accompanying text.

these tools raise draws from the long literature on settlement bargaining, emphasizing the idea that litigation's "haves" will have more precise information about the likely outcome of the case and the best arguments to lay before the judge to get there. If true, will superior information yield settlement bargaining power over those with less?

This is a harder question than it might seem at first blush. Standard Coasean models of litigation and settlement are of little use because they assume that parties have common knowledge of one another's beliefs about who would win if the case does not settle.[276] This does not describe one-sided use of legal tech, whose very purpose is to improve the (paying) party's information.[277] More apposite are a family of models that involve one-sided asymmetric information and some form of equilibrium bargaining.[278] Taken

---

[276] For an overview of settlement theory, see Andrew F. Daughety & Jennifer F. Reinganum, *Settlement*, *in* 8 ENCYCLOPEDIA OF LAW AND ECONOMICS 386, 386-71 (Chris W. Sanchirico ed., 2d ed. 2012).

[277] Unless one can construct a mechanism through which legal tech output would always credibly be conveyed to the other side, the Coasean common knowledge assumption is tough to defend. Of course, the side using legal tech could just show the other side printed output. But that might only sometimes be in the (tech-using) defendant's interests. In cases in which the defendant did not display results, it might not be discernible to the plaintiff whether the defendant actually used tech and just got a result that would improve the plaintiff's bargaining situation. There is also the possibility of unraveling, as described by Steven Shavell, *Sharing of Information Prior to Settlement or Litigation*, 20 RAND J. ECON. 183, 188 n.11 (1989), in the general discovery context. That said, if Coasean models are the right ones, then legal tech's effect might be ambiguous. For more on patterns of settlement and litigation in a "reduced form" model of litigation, see generally Jonah B. Gelbach, *The Reduced Form of Litigation Models*, 71 J.L. & ECON. 125 (2018). For a first-of-its-kind effort to model settlement outcomes in the presence of outcome-prediction tools, though focused on a situation in which all participants, both litigants and judge, have access to predictions, see Anthony J. Casey & Anthony Niblett, *Will Robot Judges Change Litigation and Settlement Outcomes? A First Look at the Algorithmic Replication of Prior Cases* (Aug. 14, 2020), https://law.mit.edu/pub/willrobotjudgeschangelitigationandsettlement outcomes/release/1 [https://perma.cc/FKV4-7FB2]. Casey & Niblett give examples in which judges don't have access to algorithmic predictions, but both parties do, in which case settlement may either increase or fall (they describe the latter case as "atypical"). *Id.* When parties and judges all have access to algorithm predictions, five out of six versions of Casey & Niblett's model have parties with identical information and expectations about who would win in the event of litigation, so 100% of cases will settle. *Id.* (Based on our inference from the general discussion in 3.6.2, it's possible that this result will not occur in the sixth version of their model. *Id.*) If real-world litigation were characterized by such extreme circumstances, one would expect the litigation process to unravel down to the demand-letter moment, with no cases filed at all in equilibrium. Casey & Niblett, however, do not have a demand-letter stage, i.e., plaintiffs have to file suit to have a chance to receive a settlement. Accordingly, this otherwise interesting model is somewhat limited in its utility for understanding how algorithmic predictions might affect the nuts-and-bolts of pre-trial procedure.

[278] *See, e.g.*, Lucian Ayre Bebchuk, *Litigation and settlement under imperfect information*, 15 RAND J. ECON. 404, 414 (1984) (showing how the informational asymmetry between the parties could influence settlement decisions or lead to a failure to settle); Klerman & Lee, *supra*, note 252, at 211-12 (showing that under asymmetric information models, the proportion of plaintiff victories varies in predictable fashion with the legal standard, legal decision makers, and case characteristics); Ivan P'ng, *Litigation, Liability, and Incentives for Care*, 34 J. PUB. ECON. 61, 62-63 (1987) (discussing the role of asymmetric information in settlement decisions); Jennifer Reinganum & Louise L. Wilde, *Settlement, Litigation, and the Allocation of Litigation Costs*, 17 RAND J. ECON. 557, 561-62 (1986) (describing how information asymmetry about damages can affect settlement, even when there is symmetrical information about the probability of judgment).

at face value, these models contemplate at least some information-sharing in equilibrium, making them potentially useful for thinking about how changes in one side's information might affect party outcomes. Distilled to their essentials and glossing over substantial complexity, these models suggest that defendants armed with superior information will enjoy better litigation and settlement outcomes than less informed plaintiffs. The reason is that, without precise probabilities, defendants facing a slew of suits cannot tell the stronger cases from the weaker ones and so must settle at a weighted average of their probabilities. With better information about the probability of a win in each case, defendants can litigate the weak cases and settle the strong ones.[279] More

---

[279] For an example, consider the Bebchuk screening model with informed defendants. *See* Bebchuk, *supra* note 278, at 406-07. In this model, there are many cases, and there is a distribution over the plaintiff's probability of winning in the event of trial. Thus, plaintiffs are highly likely to win some cases and less likely to win others. The plaintiff makes a settlement demand, and the defendant either accepts or rejects. If the defendant rejects, the case goes to trial. In each case, the defendant knows the probability with which the plaintiff will win, but the plaintiff knows only the overall distribution of probabilities with which plaintiffs win. Thus, plaintiffs must choose their settlement demand behind a veil of ignorance about their probability of winning. Defendants facing plaintiffs with strong cases will accept the settlement demand, while those facing plaintiffs with weak cases will go to trial.

To address the possibility that legal tech allows defendants to usefully refine their beliefs about plaintiffs' win probabilities, consider a set of cases in which the plaintiff's probability of winning is P; call these, "P-type cases." Suppose that among P-type cases, some are actually the sub-type in which plaintiffs would win with probability $P_{low} < P$, and some are the sub-type in which plaintiffs would win with probability $P_{high} > P$. We assume that without legal tech, at least some defendants in P-type cases can't tell the difference between P_{low} and P_{high} cases. For these defendants, P equals a weighted average of P_{low} and P_{high} (the weights are the shares of cases that are of the respective type). Defendants who use legal tech can always tell the difference. (We allow that there may be some defendants who know as much without as with legal tech. If these defendants are arrayed at the extremes of the type distribution, then it should be possible to construct the model such that the overall distribution of plaintiff win-probability types will be unaffected by the introduction of legal tech, which simplifies the rest of our discussion.) Thus, in our simple extension of the screening model, legal tech allows defendants to refine their knowledge of what would occur if the case went to trial. Note that the issue of whether credible voluntary disclosure is possible or desirable arises here. In the screening model with informed defendants, defendants benefit from being informed, because they get to litigate only when the plaintiff is weak. Thus, it might not be in their interests to share information with plaintiffs. We will assume for simplicity that no information sharing occurs.

Assuming that the overall distribution of case types is the same with legal tech as without, plaintiffs have the same information as before and thus make the same settlement offers. Now suppose P_{low} is low enough that defendants would litigate a case with that probability of plaintiff's win, and let P be high enough that defendants would not litigate a case with that probability. Without legal tech, defendants would settle all P-type cases. With legal tech, defendants will choose to litigate those P-type cases that have probability P_{low} of plaintiff win. Plaintiffs are worse off as a result of legal tech, because they now have to litigate a weak case that previously would have settled for an amount pegged in part to the settlement value of stronger cases. What about defendants in P-type cases that have probability P_{high} of a plaintiff win? Defendants using legal tech will continue to settle these cases, because $P_{high} > P$, and we know from the no-legal-tech world that P-type cases are best settled from defendants' point of view. Thus, plaintiffs in P_{high} cases are unaffected by the adoption of legal tech. In other words, unilateral adoption of legal tech by defendants makes some defendants better off at the expense of their corresponding plaintiffs, and leaves all other parties unaffected.

research is plainly needed to say something systematic on this point, but our intuition—concededly contestable and not founded on a single theoretical framework or rigorous empirical test—is that, on balance, unilateral use of legal tech can be expected to benefit the party using it, while harming less-informed opposing parties.

In the e-discovery context, where legal tech confers clearer advantages, a growing literature proposes ways to mitigate distributive concerns. Among the fixes are creation of an ethical duty to disclose defects in the other side's TAR protocol,[280] or the subjection of discovery-centered expert battles to *Daubert* constraints and Federal Rule of Evidence 702 in order to narrow expertise asymmetries.[281] Another proposal, as noted previously, would re-allocate seed set construction and model tuning to the requesting party—referred to as a "task allocation" rule, to distinguish it from a "cost allocation" rule—as a way to mitigate the cost-externalization and cross-party-agency concerns that afflict the system.[282] Each of these can be thought of as a discovery-specific patch on the distributive concerns raised by the continued proliferation of legal tech.

Our central claim in what follows is that, even if one or more of these silo-specific fixes could work, then legal tech's continued diffusion throughout the litigation system will place increasing pressure on, and often come to be analyzed through the lens of, a cross-cutting and critically important tenet of the adversarial system: the work product doctrine. In particular, if legal tech is unevenly distributed and is seen to confer a significant advantage, then litigants will seek the other side's machine outputs. What labels did you apply to the seed set? How much does your software say this case is worth? What legal arguments did your software say would be most persuasive? Faced with these questions, judges will increasingly be asked to decide whether and when the venerable work product rule should bend.

---

[280] *See* Remus, *supra* note 34, at 1716 (describing the Sedona proposal to create a Rule 3.4 violation for failure to suggest a revised predictive coding protocol that captures documents known to be responsive); *see also id.* at 1715 (advocating for a broader duty to ensure that opposing party has access to needed technology); Endo, *supra* note 34, at 863 (same).

[281] Note that this could raise barriers to entry to participate in discovery disputes at all. As it is, Federal Rule of Evidence 702 (and *Daubert*) typically do not apply to the pre-trial stage—though a growing chorus argues that it should apply to predictive coding. *See, e.g.*, Daniel K. Gelb, *The Court as Gatekeeper: Preventing Unreliable Pretrial e-Discovery from Jeopardizing a Reliable Fact-Finding Process*, 83 FORDHAM L. REV. 1287, 1297 (2014) (arguing that courts should act as "gatekeepers" of e-discovery methods); David J. Waxse & Brenda Yoakum-Kriz, *Experts on Computer-Assisted Review: Why Federal Rule of Evidence 702 Should Apply to Their Use*, 52 WASHBURN L.J. 207, 220 (2013) ("[S]earch methodologies such as computer-assisted review should be treated as an expert process subject to Rule 702 and *Daubert* challenges."); *see also* Kitzer, *supra* note 117, at 215 (echoing the "gatekeeper" view of *Daubert* and FED. R. EVID. 702).

[282] *See supra* notes 166–170 and accompanying text.

2. *Hickman*'s Work Product Bargain

For better or worse, the American litigation system is a thoroughgoing adversarial one.[283] This litigant-driven system pits the parties against one another on virtually all matters, but particularly discovery, by requiring the combatants to negotiate a mutual exchange of information in order to surface all claims and defenses and the materials relevant to each. The judge is called in only to resolve disagreements that arise during an otherwise non-public process. While American law is full of paeans to this adversarial approach and the role lawyers play within it, much of the hard work of maintaining it is a quiet, technocratic corner of civil procedure: the work product doctrine. First set forth in the U.S. Supreme Court's opinion in *Hickman v. Taylor* and inserted into federal and state rules of civil procedure thereafter, the work product doctrine protects from an adversary's discovery those documents and other tangible and intangible "things" that are prepared at the direction of counsel in anticipation of litigation.[284] Importantly, though the doctrine's protection is near-absolute in cloaking attorney mental impressions and other "opinion" work product, it can give way with respect to other types of materials, dubbed "fact" work product, where the requesting party can show a compelling need.[285]

The rationale for the work product doctrine is contested, but most accounts settle upon one of two grounds. First, the work product doctrine creates a "zone of privacy" within which counsel can operate free of interference and without worry that outputs will fall into others' hands, thus permitting them to focus on zealous client representation.[286] Permitting discovery of litigation-related materials, on this view, would lead to inadequate strategic preparation and recording of information.[287] Much

---

[283] *See* ROBERT A. KAGAN, ADVERSARIAL LEGALISM: THE AMERICAN WAY OF LAW xi (2001) (describing "Adversarial legalism['s]" quick rise to prominence as a foundation of the American legal system after 2000).

[284] 329 U.S. 495, 510 (1947).

[285] FED. R. CIV. P. 26(b)(3)(A)(ii), 26(b)(3)(B).

[286] *See Hickman*, 329 U.S. at 510-11 ("[I]t is essential that a lawyer work with a certain degree of privacy, free from unnecessary intrusion by opposing parties and their counsel.").

[287] *See* Jeff A. Anderson, Gena E. Cadieux, George E. Hays, Michael B. Hingerty & Richard J. Kaplan, *The Work Product Doctrine*, 68 CORNELL L. REV. 760, 785 (1983) (voicing conventional view that work product permits attorneys to develop facts and legal theories in private, and thus more fully). For other analyses of work product's incentive scheme, see Ronald J. Allen, Mark F. Grady, Daniel D. Polsby & Michael S. Yashko, *A Positive Theory of the Attorney-Client Privilege and the Work Product Doctrine*, 19 J. LEGAL STUD. 359, 359-60 (1990), which describe the tension between the openness underlying discovery rules and the need for confidentiality, Frank H. Easterbrook, *Insider Trading, Secret Agents, Evidentiary Privileges, and the Production of Information*, 1981 SUP. CT. REV. 309, 313 (1981), who notes that allowing free-flowing information maximizes the wealth of both users but also reduces the ability of those who create the information to receive the benefits of their labor, James A. Gardner, *Agency Problems in the Law of Attorney-Client Privilege: Privilege and "Work Product" Under Open Discovery (pt. 2)*, 42 U. DET. L.J. 253, 270 (1965), who outlines the various ways

lawyerly judgment would remain "unwritten," as Justice Murphy put it in *Hickman*, depriving the system of sustained and rigorous consideration of legal obligations and options for compliance.[288] Attorneys who fear discovery of their outputs might also minimize the negative aspects and exaggerate the positive aspects of their cases. This will engender mutual (and undue) optimism among clients and even the lawyers themselves that can stymie settlement efforts and yield inefficient resort to full-blown trials.[289]

Second, the work product rule protects against free-riding on the other side's diligence. A "learned profession," as Justice Jackson famously but somewhat cryptically put it in his *Hickman* concurrence, should not be made "to perform its functions either without wits or on wits borrowed from the adversary."[290] Part of this is properly read as just a clarifying extension of the zone-of-privacy rationale: Attorneys may incompletely prepare their cases for fear of developing adverse information in the process of investigation and analysis and may even forego inquiry that might expose information helpful to the other side.[291]

But Justice Jackson's invocation of a "learned profession" and "borrowed wits" should not be read to merely restate the notion that key information will remain unwritten or that lawyers will over-memorialize case strengths and under-memorialize weaknesses. The choice of language is deliberate and embodies a second, and deeper, rationale: The work product doctrine creates the conditions necessary for a well-functioning adversarial system by safeguarding returns on, and thus investment in, legal talent. Viewed this way, free-rider constraints, lawyer privacy zones, and even the maintenance of a market for legal talent are not ends unto themselves. Rather, they are means to an ultimate and more normatively satisfying end: a legal profession with

---

in which unlimited discovery would disturb the adversarial system, and Kathleen Waits, *Work Product Protection for Witness Statements: Time for Abolition*, 1985 WIS. L. REV. 305, 327-36 (1985, who argues against the assumption the removing the work-product protection would lead to a "parade of horribles" by removing an incentive to fully investigate.

288 *Hickman*, 329 U.S. at 495.

289 Edward H. Cooper, *Work Product of the Rulesmakers*, 53 MINN. L. REV. 1269, 1283 (1969) (arguing that "lawyers would quickly become accustomed to formulation of only the most glowing prospects for success," yielding "unduly optimistic forecasts" that would inflate client expectations and undermine reasonable settlements); *see also* Waits, *supra* note 287, at 333-35 (defending work product for witness statements based on fear that discovery would lead to inaccurate recording).

290 *Hickman*, 329 U.S. at 516 (Jackson, J., concurring).

291 *See* Anderson et. al., *supra* note 287, at 785 (arguing that without work product protection an attorney may be deterred from conducting thorough research out of fear that their opponent would benefit more than their client); Cooper, *supra* note 289, at 1279 ("[A] party who did investigate would be fearful of developing potentially adverse information only to have to hand it to his opponent."); Leland L. Tolman, *Developments in the Law—Discovery*, 74 HARV. L. REV. 940, 1029 (1961) (describing how unlimited discovery could undermine the adversary system); *see also* CAL. CIV. PROC. CODE § 2016(g) (West 1983) (repealed 2005) (observing work product purpose is to encourage attorneys to prepare thoroughly and investigate favorable and unfavorable aspects of cases); OHIO R. CIV. P. 26(A) (same).

the skill, information, and professional authority necessary to counsel compliance, and accurately determine non-compliance, in an increasingly dense legal and regulatory system.[292] The work product rule, then, is the cornerstone of a deeply adversarial model of law rooted in a set of assumptions about the self-perpetuating virtues of competition—for the maintenance of lawyers' status, for the system's truth-seeking capacity, and for optimizing law compliance in a complicated world.

As with any foundational framework, the work product doctrine has not been immune from criticism. Some contend that incentives for preparation that lawyers face are so strong, and the risks of non-preparation so grave, that they will prepare regardless.[293] An edgier criticism holds that maximal preparation may not be socially optimal in the first place, and so abolishing work product might just free up resources that could be better put toward social projects other than adjudicating disputes.[294]

Sitting atop these assorted concerns, however, is a further critique of the work product doctrine—or, perhaps better put, a compromise baked into its terms from the start. Put simply, some litigants can afford better lawyers than others. Some litigants, it follows, will enjoy better counsel in understanding their legal obligations and their optimal level of compliance in a growing regulatory state. And, in turn, some litigants will enjoy a decided edge in their courthouse struggles with other litigants. The New Deal Justices and Rule 26(b)(3)'s framers were not ignorant of these concerns.[295] But they nonetheless

---

[292] *See Hickman*, 329 U.S. at 514-15 (Jackson, J., concurring) ("[I]t too often is overlooked that the lawyer and the law office are indispensable parts of our administration of justice. Law-abiding people can go nowhere else to learn the ever changing and constantly multiplying rules by which they must behave and to obtain redress for their wrongs. The welfare and tone of the legal profession is therefore of prime consequence to society . . . .")

[293] *See* Kathleen Waits, *Opinion Work Product: A Critical Analysis of Current Law and a New Analytical Framework*, 73 OR. L. REV. 385, 450 (1994) (arguing that the built-in incentives to prepare witnesses exceed any downsides of helping the opponent); Elizabeth G. Thornburg, *Work Product Rejected: A Reply to Professor Allen*, 78 VA. L. REV. 957, 966 (1992) (arguing that the work product doctrine yields only a marginal increase in the incentive to investigate); Easterbrook, *supra* note 287, at 359-61 (arguing that a stronger evidentiary privilege could exacerbate the problem of overinvestment while making the outcome of cases less accurate); Elizabeth Thornburg, *Rethinking Work Product*, 77 VA. L. REV. 1515, 1528 (1991) [hereinafter Thornburg, *Rethinking*] (arguing that attorneys who rely on their opponents' research will be less successful over time and earn poor reputations); *but see* Ronald J. Allen, *Work Product Revisited: A Comment on Rethinking Work Product*, 78 VA. L. REV. 949, 951-55 (1992) (arguing against Professor Thornburg, but conceding that some investigation would nonetheless occur in the absence of the work product doctrine).

[294] *See* Easterbrook, *supra* note 287, at 359-60 ("Because the parties' investment is influenced largely by the size of the stakes rather than by the value of the case as a precedent, they may invest far too much (as society sees things) in litigation."); Thornburg, *Rethinking*, *supra* note 293, at 1550-51 ("[W]ork product immunity costs society in duplicated efforts, repeated disputes, skewed case outcomes, and overuse of attorneys.").

[295] *See, e.g.*, Luke P. Norris, *Labor and the Origins of Civil Procedure*, 92 N.Y.U. L. REV. 462, 466-70 (2017) (arguing that one of the "central aims" of the framers of the federal rules of civil procedure

bracketed the work product rule's distributive concerns—a necessary casualty in the service of maintaining a properly functioning adversarial scheme and safeguarding the competitive virtues that flow from it.

Only once *Hickman* and the work product rule it inscribed in American civil procedure are framed in these terms and placed on their proper footing can one see the challenge that legal tech will pose for the adversarial system's continued legitimacy and operation. While a prior generation of commentators declared that it would be "an intolerable intrusion on the bargaining process to allow one party to take advantage of the other's assessment of his prospects for victory and an acceptable settlement figure,"[296] rapid advances in legal technologies and their asymmetric deployment could lead in a very different direction. If some litigants have access to legal tech's fruits while others do not, the burning question courts will increasingly face is, to invoke the pun one last time, whether the civil procedure rules should treat "borrowed bits" the same way it treats "borrowed wits."

3. Work Product for a Digital Age

Return to our two core examples: discovery battles around TAR and use of legal analytics tools to inform a party's settlement calculus and litigation strategy. How does, or should, the work product rule apply? Can a litigant, particularly a resource-strapped one who lacks access to the full legal tech toolkit or needed data, successfully demand the other side's machine outputs?

Start with the question whether a party can or should be made to share a seed set used to train a TAR model—a question, we noted previously, that has divided federal courts.[297] Such a request might aim to allow a requesting party to gauge the comprehensiveness of the responding party's production. Or, as discussed previously, it might come in response to a judge's order authorizing the requesting party to attach her own labels to a seed or training set or even to perform some or all of the work of training the model.

On a first pass through the work product rule, one might conclude that a seed set is off-limits because it is generated through counsel's judgment and skill and, more damningly, it may reflect counsel's litigation strategy. At least one court has decided as much in the context of an *in camera* letter demanded by the court and then sought by the other side.[298] A smattering of other courts

---

was to provide "countervailing power" to the less powerful and "ensur[e] that procedural rules did not reflect and magnify the economic power imbalances immanent in industrial capitalism").

296   Cooper, *supra* note 289, at 1283.

297   *See supra* notes 205–210 and accompanying text.

298   Winfield v. City of New York, No. 15-05236, 2017 WL 5664852, at *12 (S.D.N.Y. Nov. 27, 2017).

have rejected work product claims.[299] As work product claims mount in the TAR context, operational details and the analogies they inspire will matter. Some courts are apt to liken seed sets to the finite lists of key documents and witnesses that an attorney might create to prep a witness for deposition, the situation in the leading case of *Sporck v. Peil*.[300] This "process of selection and distillation," many courts have concluded, can reveal attorney mental impressions and understandings of the case and so justifies fuller, unyielding protection as "opinion" work product.[301] In the TAR context, the *Sporck* analogy might be especially strong with seed sets created via "judgmental" sampling based on counsel's weighting of particular issues or custodians, but not seed sets created using random sampling and thus drawn from the full universe of discoverable materials.[302] Yet a court could also see even seed sets of the judgmental sort as closer to the instruction manuals used to guide document review teams to ensure a form of inter-coder reliability, where the answer is less clear,[303] or liken them to a large cache of documents taken during a document inspection. Some courts have held these do not pose a risk of conveying counsel's mental impressions or revealing other strategically

---

299 *See* Hinterberger v. Catholic Health Sys., Inc., No. 08-380, 2013 WL 2250591, at *22 (W.D.N.Y. May 21, 2013) ("[N]either the scanning nor objective coding work required any access or need for such confidential information, including attorney work product.") Meanwhile, a trio of decisions from West Virginia courts found that search terms used within a predictive coding scheme did not constitute work product. *See* Burd v. Ford Motor Co., No. 13-20976, 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., No. 13-14207, 2015 WL 4137847, at *10 (S.D. W. Va. July 8, 2015) (holding that search terms used by custodians can be disclosed without revealing substance of discussions with counsel); Johnson v. Ford Motor Co., No. 13-06529, 2015 WL 4137707, at *10 (S.D. W. Va. July 8, 2015) (finding that search terms and custodian names are not work product and collecting cases), *objections sustained in part and overruled in part*, No. 13-6529, 2015 WL 6758234 (S.D. W. Va. Nov. 5, 2015).

300 759 F.2d 312, 316 (3rd 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").

301 *See, e.g.*, Shelton v. Am. Motors Corp., 805 F.2d 1323, 1328-29 (8th Cir. 1986) (holding document selection protected 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 opinion work product). Scholars have also joined this discussion. *See, e.g.*, Kitzer, *supra* note 117, at 210 (arguing that seed sets may fall under the work product doctrine); Sean Grammel, Comment, *Protecting Search Terms as Opinion Work Product: Applying the Work Product Doctrine to Electronic Discovery*, 161 U. PA. L. REV. 2063, 2066 (2013) (arguing that search terms deserve protection as work product).

302 *See* John M. Facciola & Philip J. Favro, *Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled to Work Product Protection*, 8 FED. CTS. L. REV. 1, 9 (2015) (arguing that only the seed sets developed through judgmental sampling should merit work product protection); Christian, *supra* note 34, at 511-18 (arguing that judgmental seed sets are analogous to search terms).

303 *See* Kitzer, *supra* note 117, at 211("[T]eaching the predictive coding software to identify relevant documents is indistinguishable from teaching contract attorneys to do the same by using an instruction manual or examples of relevant documents.").

valuable information because of the sheer amount of material.[304] In the TAR context, a court might even go so far as to shunt a dispute over seed sets into the separate realm of "discovery about discovery," where the usual discovery rules, including work product, are relaxed when addressing an assertion reasonably questioning another party's discovery compliance.[305]

Even less obvious is whether the work product rule applies to legal tech tools beyond discovery, such as those that predict case outcomes. For starters, machine outputs attaching probabilities to different case outcomes need not take documentary or "tangible" forms, but rather can be requested via interrogatories, thus pushing a court's inquiry into the common-law realm of *Hickman*, or perhaps Rule 26(b)(4)'s provision regarding non-testifying experts, but perhaps not *Hickman*'s partial codification in Rule 26(b)(3).[306]

More fundamentally, a legal analytics tool that requires no more than that counsel feed in the pleadings and papers to date, or a tool primed by inputting only a set of rote case facts, does not involve substantial lawyerly judgment or effort, at least for the particular litigation in question. At best, such a tool may qualify only for Rule 26(b)(3)(A)'s lower, qualified protection reserved for "fact" work product.[307] Where that threshold determination has been made, a party must show both "substantial need" for the information and "undue hardship" in obtaining its equivalent elsewhere. While courts carefully scrutinize claims of inconvenience and resource constraints in judging

---

[304] *See* Disability Rts. Council of Greater Wash. v. Wash. Metro. Transit Auth., 242 F.R.D. 139, 141-44 (D.D.C. 2007) ("[W]ith the number of those documents said to be totaling into the thousands, it would be difficult to conceive that Plaintiffs' 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."). *See generally* Christian, *supra* note 34, at 516 (discussing how the size of the document production in question may affect the judicial calculus in determining the applicability of the work product rule).

[305] For an overview of this "process-directed discovery" and arguments for and against treating it separately, see Craig B. Shaffer, *Deconstructing "Discovery About Discovery"*, 19 SEDONA CONF. J. 215, 220 (2018).

[306] For old and new commentary on the rules governing tangible and intangible work product, see Kevin Clermont, *Surveying Work Product*, 68 CORNELL L. REV. 755, 757 (1983), and Michael A. Blaise, *The Uncertain Foundation of Work Product*, 67 DEPAUL L. REV. 35, 54 (2017).

[307] This is important because many courts hold that "opinion" work product is never discoverable, elevating the work-product doctrine to something approaching an absolute privilege. *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.").

"hardship,"[308] a party can readily make out the necessary showing where a deep-pocketed, repeat-player litigant enjoys privileged access to data, making replication of an analysis more of a factual impossibility[309] than a situation raising thornier questions about the amount of expense[310] or the parties' relative resources.[311] The "need" showing, however, may prove more of a sticking point, at least in the short-term. While the implementations vary, courts typically require that the evidence in question be "essential" or "crucial" to the moving party's case.[312] This is a plastic requirement for sure, but also one that would need to stretch considerably to include machine outputs, at least given the current state of the technology.[313] Faced with these complexities, some courts have split the difference on both "hardship" and "need" by requiring cost-sharing. In one particularly apposite case, a constitutional challenge to the City of Chicago's practice of making custodial arrests even for fine-only violations,

---

[308] A standard formulation is that "hardship" requires a demonstration that it is "significantly more difficult, time-consuming or expensive to obtain the information from another source than from factual work product of the objecting party." Gucci Am., Inc. v. Guess?, Inc., 271 F.R.D. 58, 81 (S.D.N.Y. 2010).

[309] *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 Dep't Stores, Inc., No. 11-3396, 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).

[310] *See, e.g.*, 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 replication of a report was within the plaintiff's means); *In re* Experian Data Breach Litig., No. 15-01592, 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 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"). The most "have not"-friendly approach compares the cost of reproduction to the cost of securing the materials from the opposing party. *See* Sec. & Exchange Comm'n v. Thrasher, No. 92-6987, 1995 WL 46681 (S.D.N.Y. Feb. 7, 1995).

[311] *See, e.g.*, 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, technology, staffing, and expertise" in compelling production).

[312] Many courts interpret this to require that the information the party seeks is "an essential element" in its case, or of "great probative value on contested issues," that cannot be contained elsewhere. Fletcher v. Union Pac. R.R., 194 F.R.D. 666, 671 (S.D. Cal. 2000); Nat'l Cong. for Puerto Rican Rts. v. City of N.Y., 194 F.R.D. 105, 110 (S.D.N.Y. 2000). Importantly, "need" is typically separate from hardship and keyed to the importance of the information to the party's prosecution of her case and, in particular, required evidentiary showings. *See, e.g.*, Local 703 v. Regions Fin. Corp., No. 12-1561, 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") (internal citation omitted). This makes application to machine outputs an awkward one. 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 adduced at trial.

[313] *See supra* notes 66–97 and accompanying text (surveying the current limits of NLP-based legal tech tools, particularly those that perform outcome prediction).

the court compelled plaintiffs' production of an arrest database that was not yet the basis of expert testimony or evidence, but conditioned its disclosure on the City paying half the cost of its compilation.[314]

Rule 26(b)(3)'s terms bring still other complexities. Even in cases where need or hardship cannot be shown, it is possible that advanced legal analytics tools will not qualify for work-product protection in the first place if found not to have been created "in anticipation of litigation."[315] Legal analytics tools that predict case outcomes might involve substantial attorney effort during their development—including months or even years of intense, lawyerly effort to manually construct computationally useable legal ontologies and labeling data[316]—but little to no effort in their subsequent deployment beyond inputting pleadings and papers and a keystroke. The real work of developing legal ontologies and training and tuning ML models happens miles upstream, far removed from the particular case. True, *FTC v. Grolier*, one of the Supreme Court's rare explorations of work product since *Hickman*, squarely held that work product protection extends beyond the specific litigation for which the materials were prepared.[317] The question is how far. Temporal proximity did not matter in *Grolier*. As Justice Brennan's concurrence explained, materials related to hundreds or even thousands of "essentially similar" enforcement actions brought by government agencies, or the stream of cases of the "commonly recurring type" facing private sector insurers, manufacturers, and employers, could still reveal mental processes and tactical approaches relevant to current actions long after those litigations have ended.[318] However, a legal analytics tool feels distant from a particular case in more than just a temporal sense. While concededly designed for no other purpose than to counsel clients in litigation, such a tool will have been

---

314   Portis v. City of Chi., No. 02-3139, 2004 WL 1535854, at *6 (N.D. Ill. July 7, 2004).

315   The general rule is that a document was prepared "in anticipation of litigation" if it prepared or obtained "because of" the prospect of litigation. However, courts vary in the stringency with which they apply that rule. *Cf.* 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 (1998) ("[T]he lawyer must at least 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 product protection only where the "primary motivating purpose behind the creation of the document was to aid in possible future litigation").

316   For discussion, see *supra* notes 66–97 and accompanying text.

317   462 U.S. 19 (1983). The other notable work product case is *United States v. Nobles*, 422 U.S. 225, 238-39 (1975), which made clear that *Hickman*'s common-law protection of work product extends to non-attorneys performing work for attorneys, one of several ambiguities that emerged following *Hickman*. Blaise, *supra* note 306, at 54.

318   *Grolier*, 462 U.S. at 30-31 (Brennan, J., concurring).

created neither in the midst of nor in anticipation of any particular litigation, but rather for use in future litigations in only the most general sense.[319]

Moving outside Rule 26(b)(3)'s text, judicial willingness to narrow the work product rule's ambit may stem from the fact that legal tech tools fit awkwardly with several of the rule's key normative underpinnings. For instance, legal tech tools pose little risk that compelled disclosure will cause counsel to shade outputs—that is, over-emphasizing the positive, or underemphasizing the negative—that is central to the "zone of privacy" view. Similarly, concerns about free-riding, "borrowed wits," and the need to maintain a market for legal talent have little purchase when it comes to software investments. Indeed, one could argue just the opposite: a rule that rewards technological investments may do as much to shrink the market for legal talent, at least of the human variety, as it does to bolster it. Finally, there is the fact, noted at length previously,[320] that legal tech tools do not principally perform, at least for the moment, higher-order legal cognitions. Many legal tech tools are about jockeying for advantage in ways that sit outside the conventional core of litigation judgment, attorney-client communication, and law compliance. A purely machine output that compares the likelihood of prevailing in forum X as opposed to forum Y based mostly on "external" factors—including docket loads, or the political and ideological predispositions of judge and jury—is not likely to strike judges as the kind of information production that the work product rule is designed to promote. Commentators have long called for pruning of the work product doctrine—removing business advice[321] or compliance[322] from its ambit. As distributive concerns mount, legal tech tools may create similar pressure.

The above analysis is, at best, a cursory mapping of some possible fault lines within a sprawling doctrinal landscape. More thinking is needed. This is particular so because, among the three case studies offered herein, the future of the work product doctrine is plainly the wildcard, both because of Rule 26(b)(3)'s

---

[319] *Cf.* Prater v. Consol. Rail Corp., 272 F. Supp. 2d 706, 708 (N.D. Ohio 2003) (holding that a study of employee repetitive stress complaints, though performed at counsel's direction, were business and not legal work, despite being motivated by past lawsuits and risk of future lawsuits). Another analogous case is *United States v. Textron Inc. & Subsidiaries*, 577 F.3d 21, 25-26 (1st Cir. 2009), in which the en banc First Circuit faced a work product claim by a company against an IRS effort to obtain the company's analysis of the "hazards of litigation percentages," a calculation of probabilities that its tax positions would prevail if challenged by the IRS used by auditors to ensure sufficient reserves have been set aside. The First Circuit rejected the work product claim, finding that the documents were not created "because of" litigation but rather would have been created in the ordinary course of its compliance with auditor and securities filing requirements.

[320] *See supra* notes 66–97 and accompanying text.

[321] *See, e.g.*, Michele DeStefano Beardslee, *Taking the Business Out of Work Product*, 79 FORDHAM L. REV. 1869, 1874 (2011).

[322] *See, e.g.*, Christine Parker, *Lawyer Deregulation via Business Deregulation: Compliance Professionalism and Legal Professionalism*, 6 INT'L J. LEGAL PROF. 175, 188 (1999).

interpretive uncertainties and also because of its capacity to reshape large swathes of the adversarial system well beyond TAR-centered discovery disputes. Bending the work product doctrine to meet a world pervaded by legal tech will also yield significant costs that commentators have only begun to identify. Judicially compelled sharing of machine inputs and outputs, whether in the TAR context or beyond, might blunt adversarialism's inequities, but, in so doing, it could also "disable[] lawyers from providing strong and effective client representation" and weaken the many protections adversarialism affords.[323]

The challenge for courts—and, in time, rulemakers and legislators—will be how to balance these concerns and to do so under a set of procedural rules crafted and elaborated in a very different, analog era. In Part III, we step back from the case studies and, working across them, ask some wider-aperture questions that judges and policymakers will need to ask as they oversee that process and help chart the future course of American adversarialism.

III. LEGAL TECH AND "OUR ADVERSARIALISM"

Among the legal systems of the world, the American system has long been thought exceptional in its commitment to a lawyer-dominated, adversarial process. Indeed, a rich academic literature details the ways American adversarialism departs from the judge-centered approach that prevails in much of the world,[324] debates why and when the American commitment to adversary over judicial control took root,[325] and tallies adversarialism's virtues and vices.[326] But the overwhelming focus of those inquiries has been the past and present of American litigation. Legal tech's advance provides an occasion to ask different, future-looking questions: How, if at all, will American adversarialism bend in a newly digitized civil justice system? And what role will judges—and, in time, rulemakers and legislators—play in that process? In this concluding Part, we work outward from Part II's more bounded case studies and, ranging across the full legal tech toolkit, offer some concluding thoughts on these vital questions.

As before, it is important to acknowledge the limits of our inquiry. Our observations about legal tech and the future of American litigation are subject to the same caveats, noted previously, about the contingency of technological innovation. Predicting legal tech's technical trajectory is hard enough.

---

[323] Remus, *supra* note 34, at 1715, 1717-18; *see also* Endo, *supra* note 34, at 859 (examining the protections built into the adversarial system).

[324] *See* John H. Langbein, *The German Advantage in Civil Procedure*, 52 U. CHI. L. REV. 823, 843-64 (1985) (discussing the comparative benefits of active judging).

[325] *See* AMALIA D. KESSLER, INVENTING AMERICAN EXCEPTIONALISM: THE ORIGINS OF AMERICAN ADVERSARIAL LEGAL CULTURE, 1800-1877, at 151-99 (2017).

[326] *See generally* KAGAN, *supra* note 283 (offering a book-length argument about the merits and demerits of an adversarial approach).

Predicting its effects on a sprawling litigation system verges on foolhardy. Moreover, our focus on higher-tech and litigation-focused legal tech applications plainly excludes potentially important tools, among them lower-tech ones that value recurrent types of claims (e.g., personal injury torts) or online legal advice and DIY dispute resolution tools that empower litigants to go it alone or avoid formal adjudication entirely. As already noted, these applications are both important in their own right and can also shape the formal litigation system by shrinking its domain and creating pressure to adapt in response. We leave it to others to speculate about legal tech's effects beyond the formal court settings that have been the focus of our inquiry.[327]

With that established, we highlight two synthetic insights that emerge from Part II's case studies about legal tech's incorporation into the civil justice system. The first insight concerns what we think will be increasing entwinement of intellectual property and civil procedure. Section III.A engages the very different analytical foundations of IP and civil procedure, offering our rough guess at how the tensions between them will shape and be shaped by procedural innovation in a digitized litigation system. The second insight engages the competing arguments for adversarial and inquisitorial procedural models. Section III.B thus provides a brief—surely too-brief—reassessment of the German advantage John Langbein famously found in civil procedure.[328] These insights offer the beginnings of some wider conceptual frames that can inform the thinking of judges, policymakers, and academics as they help pilot the process of legal tech's incorporation into the civil justice system.

A. *An IP for Civil Procedure*

An initial insight is that, as legal tech moves to the center of the litigation system, it will increasingly draw together civil procedure and a set of concerns that more conventionally sound in intellectual property. Indeed, civil procedure's gatekeepers, including judges but also, in time, rulemakers and legislators, will preside over what amounts to a shadow innovation policy that incorporates IP considerations into current civil procedure frameworks.[329]

This innovation-and-IP framing is hardly obvious at first. After all, civil procedure and IP are vastly different. Civil procedure aims to organize the litigation process by balancing a set of meta-values, among them efficiency,

---

[327] For an overview of the burgeoning field of computer science and access to justice, see NAT'L SCI. FOUND., COMPUTING, DATA SCIENCE AND ACCESS TO JUSTICE (2019), https://docs.google.com/document/d/1m8G-0JqZjTfdurDHkYxYp2GYUOZ9LA50BJFAlP7Fa4A/edit [https://perma.cc/R773-XEE5].

[328] *See generally* Langbein, *supra* note 324.

[329] The forms of IP most relevant to legal tech are patents, copyrights, and trade secrets. Trademarks, which are focused predominantly on consumer protection from confusion, are excluded from this analysis.

accuracy, fairness, and access. It is resolutely, if imperfectly, focused on regulating conduct within the confines of formal adjudication rather than primary conduct out in the world. IP, by contrast, seeks to reward creators of knowledge goods with temporary exclusive rights to their creations.[330] It focuses predominantly on regulating the upstream, primary conduct of creators by balancing incentives to innovate against the cost of exclusivity.[331] Aside from a generic focus on crafting optimal incentives, civil procedure and IP could not sit further apart from one another.

The coming revolution in legal tech, however, will bring increasing overlap between the two. For one, courts will now have to increasingly deal with traditional IP concerns, now in a technical area that directly affects their own functions. An example that is easy to see is the optimal discoverability of algorithmic and other software tools in litigation.[332] To date, those cases have surfaced most often in the criminal context—for instance, use of the trade secret evidentiary privilege to block disclosure of the technical guts of criminal risk assessment tools used to make bail, sentencing, and parole decisions.[333] But judges hearing civil cases will increasingly face similar questions—when, for instance, a litigant embroiled in a discovery dispute demands the source code of an adversary's proprietary TAR tool. Beyond trade secrets, courts will also surely entertain suits by producers of legal tech tools asserting infringement of patent or copyright rights—and will thus grapple with the uncertainty about software's protectability that afflicts American IP law more

---

[330] While incentivizing innovation is not the sole goal of the various IP regimes, it is a primary objective. *See* Bonito Boats, Inc. v. Thunder Craft Boats, Inc., 489 U.S. 141, 151 (1989) (implying the very purpose of patent laws is to increase the range of ideas readily available to function as the building blocks of innovation); Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 546 (1985) ("The rights conferred by copyright are designed to assure contributors to the store of knowledge a fair return for their labors."); Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 485 (1974) (referring to trade secret law as an alternative regime to patent law for the purpose of incentivizing innovation).

[331] On paper, IP does affect end users. *See* 35 U.S.C. § 271 (defining "infringement" to include "use" of patented inventions along with production and sale); 18 U.S.C. § 2319 (criminalizing moderate scale copyright infringement). In practice, however, IP rights—especially patent and trade secret protections—are most often enforced against competitors or other such "deep pockets," not end consumers.

[332] *See, e.g.*, Sonia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1274-79 (2019) (discussing the problems with protecting trade secrets while permitting discovery of source code).

[333] *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1368 (2018) (describing how some actuarial tools used to predict recidivism can be exempt from the rules of evidence); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 704 (2018) (noting that developers of criminal justice algorithms have "pursued trade secret protection" as an alternative to patents); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 117-120 (2019) (discussing the "substantial civil rights concerns that algorithms raise" in criminal justice and the barrier posed by trade secrets).

generally.[334] Decisions in each of these areas of IP law will help shape legal tech's cost structure and its distribution within the system.[335]

But these disputes, while drawing civil procedure and IP closer, will not be the sole, or even the most important, point of intersection. Indeed, Part II's case studies suggest that an equally and perhaps more important collision will center on disputes over disclosure of the inputs and outputs of legal tech tools, not in disputes over IP rights, but rather in a much wider range of litigation disputes adjudicating other types of rights. This fact is significant, for it means that civil procedure will serve as the front-line regulator of legal tech in the crucial early years of its incorporation into the civil justice system, critically shaping its use by litigants and the market for its production and distribution.[336] This is particularly so because legal tech tools derive much of their value from their exclusivity—i.e., the fact that one litigant has them and the other does not—and civil procedure rules can either bolster or undermine that exclusivity.[337] As a result, in making procedural choices, judges and policymakers will preside over what amounts to a shadow innovation policy, weighing the benefits of exclusivity against its costs.

Concrete examples abound—and were sprinkled throughout Part II's case studies. When a litigant embroiled in a discovery dispute demands disclosure

---

[334] Uncertainty about the patentability of software arises from the distinction between laws of nature, ideas, and applications. *See* Gottschalk v. Benson, 409 U.S. 63, 67 (1972) ("[A]n idea of itself is not patentable."); Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 569 U.S. 576, 580 (2013) (holding that a law of nature is not patentable); Alice Corp. Pty. v. CLS Bank Int'l, 573 U.S. 208, 216 (2014) ("Laws of nature, natural phenomena, and abstract ideas are not patentable."). In particular, the Supreme Court has held that mathematical processes are too abstract unless the invention includes an "inventive concept" via its incorporation into a real-world application. Mayo Collab. Servs. v. Prometheus Labs., Inc., 566 U.S. 66, 72-73 (2012); *Alice*, 537 U.S. at 217-18. It follows that, alone, computer algorithms are often nonpatentable. *Cf.* Bilski v. Kappos, 561 U.S. 593, 612 (2010) (rejecting a patent for a business procedure). Importantly, software's imperfect protectability fuels resort to the trade secrets evidentiary privilege. *See* Ram, *supra* note 333, at 703-04 (explaining how creators of criminal justice algorithms often seek trade secret protection).

[335] *See supra* note 147 and accompanying text.

[336] As already noted, France has already taken steps to regulate legal tech. *See supra* notes 260–263 and accompanying text. But similar legislative regulation in the United States seems unlikely. *See* Sam Skolnik, *France's Judicial Analytics Ban Unlikely to Catch on in U.S.*, BLOOMBERG L. (June 5, 2019, 12:26 PM) https://news.bloomberglaw.com/us-law-week/frances-judicial-analytics-ban-unlikely-to-catch-on-in-u-s [https://perma.cc/8J55-TFK9] (quoting legal tech experts who believe France's regulations will have limited influence in the States).

[337] The exact degree to which exclusivity will affect legal tech's value is unclear. In general, scarcity tends to drive up a good's value. *See* Michael Lynn, *Scarcity Effects on Value: A Quantitative Review of the Commodity Theory Literature*, 8 PSYCH. & MKTG. 43, 46-47 (1991) ("[S]carcity enhances the value of anything that can be possessed."). However, software products are particularly successful at overcoming the issue the scarcity principle causes. Vasilis Kostakis & Andreas Roos, *New Technologies Won't Reduce Scarcity, but Here's Something That Might*, HARV. BUS. REV. (June 1, 2018), https://hbr.org/2018/06/new-technologies-wont-reduce-scarcity-but-heres-something-that-might [https://perma.cc/W82R-HT6P]. The technology may also be so beneficial that the cost of missing out when other parties possess it may outweigh the downsides of having to share the technology with others.

of an adversary's seed set in order to contest the completeness of a document production, a trial judge must determine whether to compel sharing of the entire seed set, only positively flagged documents (which, as noted previously, are the only "relevant" ones within the meaning of Rule 26(b)(2)), or none at all. As Section II.A showed, courts are all over the map on which level of disclosure they require and in what circumstances.[338] But it is not hard to see how an accretion of rulings on the issue will determine the value of TAR tools to litigants and, by extension, the incentives for further innovation.[339] Compelled sharing of TAR inputs could depress the technology's development—or convince litigants not to use it at all, either because it confers little advantage or, worse, risks putting non-responsive and privileged documents into an adversary's hands.[340] Conversely, judicial decisions on motions to compel permitting requested discovery conditional on use of TAR, as judges have begun to do, will spur use of TAR and, with it, grow the market that produces it. Wise decisions on these questions must, whether explicitly or implicitly, weigh TAR's utility (e.g., its potential to reduce litigation costs) against its costs (e.g., adversarial inequities, discovery abuse).

Similar questions will condition the development and use of other parts of the legal tech toolkit, most notably outcome-prediction tools. As explored in Section II.B, judicial demands for the outputs of outcome prediction engines—whether in connection with motions practice around choice-of-forum or choice-of-law disputes, such as a party's request to transfer venue, or even dispositive motions seeking summary judgment—will depress the value of those tools to litigants, potentially limiting their use and slowing their development. Similarly, if judges compel adversaries to share machine outputs—for instance, by compelling party responses to contention interrogatories requesting them—the value and use of those tools, and the market for their production, could contract substantially. As legal tech tools advance in sophistication, these and other

---

338 *See supra* notes 205–208, 296–297 and accompanying text.

339 *See* Christian, *supra* note 34, at 510 (noting that judicial resolution of whether and when TAR inputs, especially seed sets, are subject to privilege could "impede the acceptability of predictive coding technology in civil litigation").

340 There are several factors at play here, some of which may push for a market's existence even in the face of judicial rulings. Some "haves" say that simply knowing what a tool says isn't sufficient for "have nots" to pose a threat. *See* Victoria Hudgins, *They Come in Peace: Why the Legal Research Market Welcomes Nonprofit Entrants*, LAW.COM (Jul. 21, 2020, 11:30 AM), https://www.law.com/legaltechnews/2020/07/21/they-come-in-peace-why-the-legal-research-market-welcomes-nonprofit-entrants [https://perma.cc/RVN3-2RRG]. In many cases, tech just tells a user where to look and what to pay attention to. Therefore, a scenario where both parties are equipped with, for example, an outcome prediction tool (or its output) would be preferable to a scenario where neither party has it. Each party still needs to know what to do with the output to maximize the utility gain. The haves are confident in their ability to make the most of the data using their better-funded resources. Granted, that still may mean that obligating disclosure of the inputs and outputs of legal tech tools may still increase the justice gap between haves and have nots.

conflicts will sharpen the tension between civil procedure values and the incentives to produce and use the tools in the first place.

In resolving these tensions, IP has much to offer civil procedure because it provides a ready-made vocabulary and a familiar set of conceptual frameworks for formalizing and weighing the trade-offs between legal tech's benefits (efficiency, accuracy) and the distributive and other costs that derive from its exclusivity. The benefits are substantial, for the many procedural doctrines implicated by legal tech do not expressly consider trade-offs between innovation incentives and the social costs of exclusive rights that are the fundamental analytic building blocks of IP. The work product doctrine, or any other part of civil procedure for that matter, was simply not built for that wider, innovation-focused inquiry. Going forward, however, judges may well incorporate the considerations first established in the IP realm into the civil procedure space. Questions of efficiency, for example, will no longer be limited in scope to litigation but will, even if only implicitly, expand to include market effects.

As legal tech proliferates through the civil justice system, additional regulatory opportunities may present themselves. Indeed, an innovation-and-IP frame helps us to imagine potential interventions other than a judge-led process of muddling through with procedural tools built for other tasks. As a growing literature in IP makes clear, trade secrecy and conventional IP protection are but two levers in a wide portfolio of innovation policies that also includes prizes, grants, and tax incentives, which can be coupled with disclosure requirements or other conditions.[341] If proliferating legal tech opens up distributive divides by allowing the "haves" to systematically win out over the "have nots," one could imagine a wide range of policies focused on mitigating distributive costs without threatening the robust market for legal tech. Indeed, it is precisely where conventional IP rights fail to yield a socially optimal outcome that prizes and public subsidies may be preferred to conventional IP rights.[342] Among the possibilities are a government-funded open-source legal tech platform designed to provide litigation's "have nots" with a baseline set of tools, or even a courthouse discovery arm, with technologist magistrates or law clerks, that oversees or even performs e-discovery.[343]

---

341   *See, e.g.*, Daniel J. Hemel & Lisa Larrimore Ouellette, *Innovation Policy Pluralism*, 128 YALE L.J. 544, 544 (2019) (stating that elements of IP protection can be combined with incentives like prizes, tax preferences, and government grants); Ram, *supra* note 333, at 700-01 (describing policy mechanisms in addition to trade secret protection that can promote innovation).

342   *See, e.g.*, Daniel J. Hemel & Lisa Larrimore Ouellette, *Beyond the Patents-Prizes Debate*, 92 TEX. L. REV. 303, 376, 381 (2013) (describing the economic benefits of these subsidies compared to the deadweight losses of patent monopolies).

343   *See* Re & Solow-Niederman, *supra* note 3, at 285 (noting the possibility of "public option" legal tech or public subsidization or provision of data as "an institutional counterweight to proprietary datasets"). For an excellent and broader introduction to the ways in which access to data is becoming an access to justice issue, see Alexander & Feizollahi, *supra* note 243.

Of course, there are limits to the IP analogy in thinking about legal tech's future. IP is primarily focused on the *production* of innovation, not its use.[344] But in the legal tech context, work product and other rules confer exclusivity on the user, not the inventor.[345] IP's critics have also long groused that IP under-incentivizes production of social benefits—e.g., environmental benefits[346]—or is too solicitous of inventions that lack any objective social value, as one might say of a legal tech tool that promotes pure rent-seeking behavior by litigants who uniquely possess it.[347]

In short, many details of an IP-and-innovation framework remain to be worked out. But the potential implications are substantial: The coming revolution in legal tech will require judges and policymakers to incorporate a new covering value into the traditional pantheon of efficiency, accuracy, fairness, and access. Innovation incentives, not just these traditional procedural values, will become a central feature of the meta-level procedural calculus.

B. *Legal Tech and the German (Dis)Advantage*

If marrying civil procedure and IP was all that was needed to oversee legal tech's incorporation into the litigation system, then the process, and the analytics required, might seem manageable. For instance, in the e-discovery

---

[344] *See supra* note 329; U.S. CONST. art. I, § 8, cl. 8 (granting Congress the power "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries"). *See generally* 1 PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE 2020 (2020) (surveying federal IP law).

[345] Despite this apparent inconsistency and the fact that regulating legal tech through the civil procedure lens would likely directly target users of legal tech, the policies would also affect the incentives of the entities creating the legal tech. In fact, the rationale for work product seems oddly similar to the innovation rationale in IP: to encourage the party producing something to put as much effort into it as they can instead of hedging against copying. In an interesting overlap between the two, "protecting creators' work product" has been cited as a motivation for copyright law. VHT, Inc. v. Zillow Grp., Inc., 918 F.3d 723, 739 (9th Cir. 2019).

[346] *See* Christopher Buccafusco & Jonathan S. Masur, *Intellectual Property Law and the Promotion of Welfare* 13 (Univ. of Chi. Pub. Law & Legal Theory, Working Paper No. 607, 2017) ("[T]he same patent rules that are helpful for pharmaceuticals might be harmful for environmental technologies."); Hemel & Ouellette, *supra* note 341, at 555-56, 575-76 (explaining that market institutions may inadequately reward environmental innovations); Hemel & Ouellette, *supra* note 342, at 328-29 (stating that patents are ineffective at linking social value and private returns).

[347] Courts do not typically require patented inventions to serve a particularly useful function. *See* Juicy Whip, Inc. v. Orange Bang, Inc., 292 F.3d 728, 745 (Fed. Cir. 2002) ("[A]n invention's deceptive nature has no bearing upon its utility . . . ."). The more modern trend is that an invention has utility so long as it carries out a function specific to the invention (specific utility) which applies to the real world (substantial utility). The invention does not need to be performed well, nor does the invention need to necessarily improve the public. The invention just cannot fail to carry out its asserted function at all. Grunenthal GmbH v. Alkem Labs. Ltd., 919 F.3d 1333, 1345 (Fed. Cir. 2019); Gene Quinn, *Understanding the Patent Law Utility Requirement*, IPWATCHDOG, https://www.ipwatchdog.com/2015/11/07/understanding-the-patent-law-utility-requirement/id=63007 [https://perma.cc/8CVM-6F9T].

context, determining when to compel party collaboration or sharing of machine outputs will turn on a tractable inquiry weighing litigation cost-savings, the equity costs of gaming and discovery abuse, and litigant incentives to use TAR in the first place.

But Part II's case studies suggest a second broad framing, and a second challenge, that is further-ranging and more complex. In particular, legal tech's integration into the civil justice system will also entail explicit or implicit judgments about the optimal distribution of information within the system, both horizontally, between litigants and litigants, and also vertically, between judges and litigants. In determining the distribution of information along these two axes, judges and policymakers will help set the balance of adversary and judicial control within the system and, in so doing, shape the future course of American adversarialism.

While there are many ways to conceptualize this framing, a good way to start is to consider, as students of American procedure long have, the contrast between American adversarialism and the more judge-centered approach that prevails in much of the rest of the world, particularly Continental Europe.[348] The so-called "German advantage" in civil procedure, coined in Langbein's iconic 1985 study, has many rich facets. But its core claim is that the judge-centered "inquisitorial" approach, in which judges oversee the pace, phasing, and substantive direction of fact- and issue-development in cases, offers a superior alternative to an American system defined by adversary control.[349] This comparison has become a central organizing framework for thinking about the optimal mix of adversary and judicial control in litigation systems. It is the dominant lens, to cite just one example, for weighing the virtues and vices of the much-debated trend in American civil procedure toward "managerial judging," whereby American judges have steadily adopted a more intrusive approach, particularly in complex litigations, by directing the pace, content, and character of litigation.[350]

The German comparison takes on new relevance in a litigation world infused with legal tech because it lays bare a core bet that underpins the design of any litigation system about the salience of litigant-litigant as against judge-litigant information asymmetries.[351] The judge-centered, inquisitorial

---

[348] Langbein, *supra* note 324.

[349] *Id.* at 847 (contrasting the "inquisitorial zeal" of the Continental system with the "truth-defeating excesses of American adversary fact-gathering [which] cause knowable facts to be obscured").

[350] Resnik, *supra* note 39; *see also* Nora Freeman Engstrom, *The Lessons of Lone Pine*, 129 YALE L.J. 2, 60 (2019) (reviewing the literature and noting managerial judging's durability). For the traditional judicial role, see Abram Chayes, *The Role of the Judge in Public Law Litigation*, 89 HARV. L. REV. 1281, 1286 (1976).

[351] Only one analysis of which we are aware connects legal tech to the choice between an adversarial and inquisitorial system, but it does so only in passing in making the argument that disillusionment with lawyer advocacy and its perceived capacity to mobilize bias and leverage

approach is premised on an intuition that judge-litigant information asymmetries will not be as significant, or as consequential, as litigant-litigant information asymmetries.[352] Judges might not know as much as the parties, but they can nonetheless oversee litigation's conduct, steering the phased and targeted acquisition of evidence and witness examination, all the while searching for the "jugular" issue that permits an early and definitive end to a case.[353] By implication, the inquisitorial approach is founded on the further view that litigant-litigant asymmetries *are* significant—and that an unregulated adversarial process in which litigants fend for themselves, and some can afford more and better counsel, will yield significant costs in efficiency and equity.[354] Judge control of the proceedings, on this view, is a hedge against adversarialism's excesses.

In stark contrast, the American system is built upon the notion that judge-litigant asymmetries are apt to be substantial, and that judicial control over the proceedings will yield too many inefficiencies and errors. Litigants know their case better than judges, and so it is only partisan fact-gathering, which aligns responsibility and incentive, that can consistently achieve a full ventilation of facts.[355] One need not deny the costs of partisan control in order to hold this view. Rather, the costs of a blindered judge running the show outweigh the social costs of the inequities created in a world in which some litigants can afford more and better counsel than others.

Framing litigation design in these terms powerfully captures the stakes of legal tech's advance. A key question going forward will be whether legal tech's proliferation throughout the civil justice system will shift the core bargain

---

ignorance might push adversarial systems in an inquisitorial direction. *See* Re & Solow-Niederman, *supra* note 3, at 275. But the authors go on to note that AI's capacity to "pull[] back the judicial curtain" could equally undermine judges and a more inquisitorial approach. *Id.*

352 Langbein, *supra* note 324, at 843 (noting that the "active role of the judge places major limits on the extent of the injury that bad lawyering can work on a litigant").

353 *Id.* at 830-31 (noting "jugular" idea and fact that a German court "functions without sequence rules" and, in particular, without any distinction between pre-trial and trial); *id.* at 846 (noting that "judicial control of sequence works to confine the scope of fact-gathering to those avenues of inquiry deemed most likely to resolve the case").

354 *Id.* at 343 ("[V]ery little in our adversary system is designed to match combatants of comparable prowess . . . [T]he active role of the judge [in Germany] places major limits on the extent of the injury that bad lawyering can work on a litigant."); *see also* William W. Schwarzer, *Managing Civil Litigation: The Trial Judge's Role*, 61 JUDICATURE 400, 405-06 (1978) (offering another classic statement of the problem). A further claim is that inquisitorialism protects litigants from their own counsel, mitigating lawyer-client agency costs. *See* E. Donald Elliott, *Managerial Judging and the Evolution of Procedure*, 53 U. CHI. L. REV. 306, 330-32 (1986) (discussing the principal-agent problem for lawyering in an adversarial system).

355 *See, e.g.*, Robert G. Bone, *Who Decides? A Critical Look at Procedural Discretion*, 28 CARDOZO L. REV. 1961, 1991 (2007) ("The conventional assumption underlying the commitment to adversarial fact-finding is that competition between adversaries is likely to ferret out the truth."). Of course, Bone contests this assumption, calling it "excessively optimistic." *Id.*

that undergirds the American commitment to adversary control, whether across the board or for particular tools or litigation types. It also allows us to imagine a set of possible futures as legal tech proliferates. Table 2 makes this concrete by representing, in stylized form, four different combinations of litigant-litigant and judge-litigant information asymmetries that might emerge over time as the market for legal tech takes shape. For each combination, one can then ask how, or if, procedure should adjust in response.

Consider first Table 2's northwest quadrant—a fully democratized system in which litigants and judges alike have access to the complete legal tech toolkit.[356] In this scenario, there are few distributive concerns, and the choice between adversary and judicial control becomes less salient, since all sides have the same degree of transparency into the pool of evidence, probabilities over case outcomes across different fora, and even the assigned judge's own (latent) predispositions in similar past cases. To be sure, this future is not unproblematic. As noted in Section I.C, some might worry that such a system will suffer from creeping automation bias—undue reliance by actors within the system on machine outputs—and drain the system of its capacity to adapt to social change or apply equitable principles in hard cases. Others might further worry that an uptick in settlements driven by a leveling of information will depress or even distort the set of litigated cases on which predictive tools can be trained and updated,[357] stymie the public elaboration of legal norms,[358] or reduce litigant conduct of socially valuable but privately costly discovery.[359] Compared to other options, however, the northwest future may provide an ideal baseline.

---

356 Note that this is the situation modeled in Casey & Niblett, *supra* note 277.

357 *Compare* Alarie et al., *supra* note 93, at 254 ("These technologies increase the likelihood of settlement, while the likelihood of cases going to court will fall, save perhaps for the most ambiguous, where further legal development will be most valuable."), *with* Casey & Niblett, *supra* note 277 (noting concern about both "a reduction in the production of judicial precedent" and also selection bias because "only cases with close and perhaps confounding factual situations end up in court," yielding a set of past cases upon which predictive tools can be trained that "may not reflect the full picture" of litigation).

358 *See supra* note 126 and accompanying text. For the classic critique of pervasive settlement, see Owen M. Fiss, *Against Settlement*, 93 YALE L.J. 1073, 1075 (1984). For a full airing of the debate, see generally Symposium, *Against Settlement: Twenty-Five Years Later*, 78 FORDHAM L. REV. 1117 (2009).

359 *See* Zambrano, *supra* note 150, at 72 (reviewing literature showing how discovery costs affect the ability of the system to produce fair outcomes).

Table 2: Legal Tech and Information Asymmetries: Four Futures

	Narrower judge-litigant information asymmetries	Wider judge-litigant information asymmetries
Narrower litigant-litigant information asymmetries	*Democratization*: Litigants and judge alike enjoy full access to legal tech outputs; system steadily approaches full and fully distributed information; significant transparency, potential uptick in pre-trial settlements.	*Bilateral Litigant Supremacy*: Litigants enjoy full and equal access to legal tech outputs; judges do not—and will (should?) recede from view.
Wider litigant-litigant information asymmetries	*Power Sharing by Judges and Litigant "Haves"*: Concentration of information in hands of "haves," but judges have tech, too, and may protect have-nots	*Unilateral Litigant Supremacy*: Information concentrated in hands of "haves," who lord over adversaries and judges alike; a litigation dystopia

Now move to Table 2's northeast quadrant, in which litigants on both sides of the "v" pervasively use legal tech tools but judges do not.[360] This system features wide information asymmetries of the vertical (judge-litigant) sort but not the horizontal (litigant-litigant) sort. As with the first scenario, we have few concerns about distributive effects resulting from an unequal distribution of technological capacity across parties. Armed with the same information about case outcomes, and with the optimism bias and information asymmetries that can stymie negotiations banished from the system, rational parties will quickly reach settlements.[361] In this world, there is little warrant for a procedural response along the lines of compelled sharing of legal tech's inputs or outputs. While one might still worry about the loss of public elaboration of legal norms or the ability of the system to update in a pool of litigated cases that may suffer from selection bias, the system will

---

[360] Note that this is the situation treated, in a brief appendix, in Casey & Niblett, *supra* note 277 (considering the situation in which "the parties have access to litigation prediction tools (but judges and other decision makers do not)"). It is also the world predicted by Eugene Volokh, who argues that legal tech is likely to come to lawyers before judges. Volokh, *supra* note 3, at 1151.

[361] *See* Alarie et al., *supra* note 93, at 233 (noting that the closing of informational asymmetries will eliminate "wasteful expenditure on litigation").

have nonetheless achieved something close to the *Hickman* ideal: "mutual knowledge of all the relevant facts."[362]

Note, however, that this does not necessarily mean preservation of the status quo as a procedural matter. In a world defined by bilateral litigant control of legal tech, there is a strong case to be made that judicial control—including the "managerial judging" that has increasingly characterized the American system—will and, indeed, *should* erode. For some, a level legal tech playing field would provide a welcome opportunity to pare back an overweening judicial presence that, as Judith Resnik cogently noted long ago, often plays out "beyond the public view, off the record, with no obligation to provide written, reasoned opinions, and out of the reach of appellate review."[363] If that view prevails, democratized but litigant-centered legal tech will likely move the American system further away from a proto-German, inquisitorial model.

Turning southward, Table 2's southwest quadrant captures a system characterized by wide horizontal but narrow vertical information asymmetries. As a practical matter, this future is harder to glimpse, and it seems least likely to unfold in reality. In particular, it would require a legal tech toolkit that is available to and adopted by budget-strapped courts but not large classes of litigants. Still, consideration of the southwest future's contours is instructive. One possibility is that such a system might generate fewer distributional concerns, since the judge will be well-positioned to level the litigant playing field, at least where a dominant litigant deploys legal tech tools for pure rent-seeking purposes. Note that the case for managerial judging here is strong. In stark contrast to the prior scenario, the answer here might be *more* managerial judging and *more* judicial control over the conduct of the proceedings.

The most concerning of all the scenarios—a kind of litigation dystopia—is the southeast quadrant, a system characterized by wide asymmetries along *both* dimensions (litigant-litigant, judge-litigant). A relatively narrow set of litigants— likely well-heeled ones—would exercise something like unilateral control over legal tech's informational advantages and could thus engage in litigation rent-seeking, using their privileged command of case outcomes to choose the most advantageous forum, game the discovery process to ensure that the most

---

[362] *Hickman v. Taylor*, 329 U.S. 495, 507 (1947). Casey and Niblett complicate this rosy picture somewhat by noting that, in certain circumstances (e.g., the litigants, rather than suffering from the usual optimism bias about their chances that the predictive tools helps to correct, instead both hold views about the plaintiff's chances that are higher than the algorithmic prediction), better information can reduce the likelihood of an efficient settlement. Casey & Niblett, *supra* note 277.

[363] Resnik, *supra* note 151, at 378, 380; *see also* Langbein, *supra* note 324, at 861 (noting the rise of managerial judging "has not been accompanied by Continental-style attention to safeguarding litigants against the dangers inherent in the greatly augmented judicial role"); Todd D. Peterson, *Restoring Structural Checks on Judicial Power in the Era of Managerial Judging*, 29 U.C. DAVIS L. REV. 41, 78 (1995) (making a similar argument and emphasizing the threat to party "autonomy"). *See generally* Engstrom, *supra* note 150, at 60 (summarizing the literature).

damaging evidence remains under wraps, and craft winning legal arguments that reflect the latent predilections of particular judges. Legal tech's "haves" would systematically win out over its "have nots," whether in pre-trial jockeying over dispositive motions at trial or in the capture of settlement surplus.

Here, the case is strongest for a procedural response, via compelled sharing of legal tech tools, judicial demands for litigants' machine outputs, or both. But note that it is here that the incentives to innovate may be at their most powerful—and one might expect a hyper-sophisticated technological trajectory, or substantial effort by litigation actors with privileged access to data to develop a suite of advantage-conferring proprietary tools. Here, the IP-and-innovation framing helps to mark out the trade-offs and even points the way to some other potential remedies. First, procedural interventions can level the playing field but only while also blunting litigant incentives to use legal tech in the first place, thus depriving the system of its potential accuracy-enhancing and cost-reducing virtues. On the other hand, the right policy decision should, in fact, be to stifle innovation if its social inequity costs are determined to outweigh the benefits rooted in improved efficiency and accuracy. Second, and as noted previously, legislators concerned that the litigation system is moving toward the southwest quadrant might consider funding a "public option" set of legal tech tools or taking substantial action to improve data accessibility as an "institutional counterweight to proprietary datasets."[364]

Of course, these four futures are unlikely to hold across the board, for all kinds of litigation, at all levels of the judicial system. A key challenge for judges and policymakers, and a key pressure given civil procedure's facial commitment to transubstantivity, will be how to craft a variegated response across litigation types to ameliorate concerns where they arise most acutely.

* * *

Looking across Part II's case studies helps us to see a larger landscape, and a wider-angle frame for thinking about legal tech's incorporation into the civil justice system. That said, neither the IP-and-innovation nor the information-asymmetry frame is comprehensively treated here. Nor do these frames exhaust the ways one could conceptualize legal tech's incorporation into the civil justice system. Neither says anything about the allocation of power as between judge and jury, another front in the procedural battles fought in recent decades, particularly around summary judgment. But taken together, they capture some of the essential puzzles that will face judges and policymakers going forward. As these actors remodel civil procedure in the years to come, they will make a wide range of judgments—about how much

---

[364] Re & Solow-Niederman, *supra* note 3, at 366.

innovation is a good or bad thing, how much exclusivity to permit, and which asymmetries and inequities to tolerate—that will help set the balance of adversary and nonadversary values within the system and, in the process, chart the future course of American litigation. The two frames thus provide a rough roadmap for the kind of work that lies ahead as the details of a newly digitized litigation system come into focus.

Conclusion

This Article has argued that legal tech is likely to reshape the American system of litigation not merely by changing how lawyers do their work, but also by resetting several of the system's procedural cornerstones. Along the way we have raised, but only partially answered, numerous questions, empirical and otherwise, that will repay research as the next era of American litigation comes into focus. We nonetheless hope that the insights offered in this Article—from our sober accounting of legal tech's likely trajectory to our case studies of tools and rules to our more synthetic thoughts on system design—can aid the thinking of judges, policymakers, and scholars as they oversee legal tech's continued move to the center of the civil justice system.

*   *   *   *   *