11-2022

Using Artificial Intelligence in the Law Review Submissions Process

Brenda M. Simon
California Western School of Law, bsimon@cwsl.edu

Using Artificial Intelligence in the Law Review Submissions Process

*Brenda M. Simon**

The use of artificial intelligence to help editors examine law review submissions may provide a way to improve an overburdened system. This Article is the first to explore the promise and pitfalls of using artificial intelligence in the law review submissions process. Technology-assisted review of submissions offers many possible benefits. It can simplify preemption checks, prevent plagiarism, detect failure to comply with formatting requirements, and identify missing citations. These efficiencies may allow editors to address serious flaws in the current selection process, including the use of heuristics that may result in discriminatory outcomes and dependence on lower-ranked journals to conduct the initial review of submissions. Although editors should not rely on a score assigned by an algorithm to decide whether to accept an article, technology-assisted review could increase the efficiency of initial screening and provide feedback to editors on their selection decisions. Uncovering potential human bias in the existing selection process may encourage editors to develop ways to minimize its harmful effects.

Despite these benefits, using artificial intelligence to streamline the submissions process raises significant concerns. Technology-assisted review may enable efficient implementation of existing biases into the selection process, rather than correcting them. Artificial intelligence systems may rely on considerations that result in discriminatory effects and negatively

 ProFlowers Professor of Internet Studies and Professor of Law, California Western School of Law; University of California, Berkeley School of Law (J.D.); University of California, Los Angeles (B.S.). I received helpful comments from participants at the BioLaw Conference at Stanford Law School and the Legal Scholars Roundtable on Artificial Intelligence at Loyola University Chicago School of Law, where I presented an earlier version of this piece. Special thanks to William Aceves, Clark Asay, Ryan Calo, Tabrez Ebrahim, Eric Goldman, Catherine Hardee, Cindy Hirsch, Margot Kaminski, Nancy Kim, Mark Lemley, Amanda Levendowski, David Levine, Nicholson Price, Sandra Rierson, Michael Risch, Matthew Sag, Joanna Sax, Ted Sichelman, Howard Strasberg, Charlotte Tschider, and Michael Wright for valuable discussions and suggestions on prior drafts.

impact groups that are not adequately represented during development. The tendency to defer to seemingly neutral and often opaque algorithms can increase the risk of adverse outcomes. With careful oversight, however, some of these concerns can be addressed. Even an imperfect system may be worth using in limited situations where the benefits substantially outweigh the potential harms. With appropriate supervision, circumscribed application, and ongoing refinement, artificial intelligence may provide a more efficient and fairer submissions experience for both editors and authors.

TABLE OF CONTENTS

INTRODUCTION

The law review submissions process has become untenable. Students serving on law reviews are often inundated with over a thousand articles each submission cycle.[1] Law review editors currently use only basic technological tools to filter the copious number of submissions received each year, such as sorting by date, title, and keywords.[2] Understandably, they sometimes rely on questionable heuristics — law school affiliation, author status, previous placements, even pressure from their own professors[3] — that can bias decision making and result in

[1] *See* Christian I. Bale, *Three Suggestions to Promote New Scholarship from an Outgoing Editor-in-Chief*, 71 DUKE L.J. ONLINE 47, 48 (2021) ("During the 2021 spring selection period, *DLJ* received 1,368 manuscripts . . . ."); Leah M. Christensen & Julie A. Oseid, *Navigating the Law Review Article Selection Process: An Empirical Study of Those with All the Power — Student Editors*, 59 S.C. L. REV. 175, 203-05 (2007) ("Several other editors from the Top 50 law schools reported that they received between 1,500 and 2,000 articles per year."); Michael J. Higdon, *Beyond the Metatheoretical: Implicit Bias in Law Review Article Selection*, 51 WAKE FOREST L. REV. 339, 341 (2016) (stating that "law reviews are reporting submission numbers as high as 2200"); *cf.* Barry Friedman, *Fixing Law Reviews*, 67 DUKE L.J. 1297, 1305 (2018) (describing the selection of articles by students).

[2] *See Law Review System*, SCHOLASTICA, https://scholasticahq.com/law-reviews (last visited July 25, 2021) [https://perma.cc/RT9U-ML57].

[3] *See* Stephen Thomson, *Letterhead Bias and the Demographics of Elite Journal Publications*, 33 HARV. J.L. & TECH. 203, 224 (2019) (finding that several of the top law reviews had at least 20 percent of their articles authored by professors at their home institutions); Albert H. Yoon, *Editorial Bias in Legal Academia*, 5 J. LEGAL ANALYSIS 309, 310 (2013) (observing that "law reviews, with few exceptions, publish a higher percentage of their articles from their own law faculty than from any other law school"). *But see* Michael Conklin, *Letterhead Bias and Blind Review: An Analysis of Prevalence and Mitigation Efforts*, 2022 U. ILL. L. REV. ONLINE 1, 9 ("The results support the notion that letterhead bias is not a significant problem in legal academia and that blind review does not significantly alter publication decisions. This research, however, only provides a singular data point that must be considered in light of the contrasting evidence that letterhead bias does exist."); Kevin M. Yamamoto, *What's in a Name? The Letterhead Impact Project*, 22 J. LEGAL STUD. EDUC. 65, 67 (2004) (finding "no significant statistical difference . . . in the time of response, type of response, or whether the article was accepted" based on letterhead, although the study was limited to the submission of a single article).

discriminatory effects.[4] Given existing disparities in academia,[5] reliance on these types of metrics can exacerbate inequality and inhibit the recognition of valuable contributions. Article placement is the currency of academia.[6] It affects almost every aspect of the academic ecosystem, from entry-level and lateral hiring to promotion and tenure, as well as speaking opportunities and compensation.[7] While anonymous review would result in a less biased selection process, many law reviews lack the capacity to devote the time and effort necessary to undertake the examination involved.[8] Proposals to address some of these concerns,

[4] *See* James Lindgren, *An Author's Manifesto*, 61 U. CHI. L. REV. 527, 530 (1994) (describing how editors screen articles based "on the prestige of the law school from which the manuscript was submitted"); Thomson, *supra* note 3, at 210-13 (discussing letterhead bias and other proxies for quality); Jonathan I. Tietz & W. Nicholson Price II, *Acknowledgments as a Window into Legal Academia*, 98 WASH. U. L. REV. 307, 343 (2020) ("Heuristics are inevitable."). *But see* Richard A. Posner, *The Future of the Student-Edited Law Review*, 47 STAN. L. REV. 1131, 1133-34 (1995) (pointing to the "reputation of the author" as a potential signal of quality).

[5] *See* MEERA E. DEO, UNEQUAL PROFESSION: RACE AND GENDER IN LEGAL ACADEMIA 3-4 (2019) (describing the "challenges and opportunities associated with race and gender that are unique to . . . underrepresented faculty"); Hannah Brenner, *Expanding the Pathways to Gender Equality in the Legal Profession*, 17 LEGAL ETHICS 261, 275 (2014) (describing "gender demographics in the legal academy"); Nancy Leong, *Discursive Disparities*, 8 FIU L. REV. 369, 373 (2013) ("Disparities . . . permeate legal education.").

[6] *See* Anthony Michael Kreis, *Picking Spinach*, 50 LOY. U. CHI. L.J. 395, 395 (2018) (describing the significance of placement).

[7] *See id.* at 395-96 (explaining that "law review articles are key in hiring, promotion, and tenure decisions"); Stephen R. Heifetz, *Efficient Matching: Reforming the Market for Law Review Articles*, 5 GEO. MASON L. REV. 629, 632 (1997) ("Decisions by law school hiring and tenure committees are often influenced by the . . . name of the law review in which the work appears.").

[8] *See* Friedman, *supra* note 1, at 1349-50 ("Review of articles ought to be blind."); Richard A. Wise, Lucy S. McGough, James W. Bowers, Douglas P. Peters, Joseph C. Miller, Heather K. Terrell, Brett Holfeld & Joe H. Neal, *Do Law Reviews Need Reform? A Survey of Law Professors, Student Editors, Attorneys, and Judges*, 59 LOY. L. REV. 1, 72-73, 75 (2013) ("The vast majority of legal professionals and student editors believe that law reviews should . . . include blind, peer reviews and more student training."); *cf.* Paul J. Heald, *The Law Review Scam: How to Humanely End Law School Exceptionalism* 3 (Univ. of Ill. Coll. of L. Legal Stud. Rsch. Paper, Paper No. 22-19, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4086728 [https://perma.cc/6U2M-PGU6] (proposing that "universities should phase in a *prospective* rule that only peer-reviewed articles count toward tenure"); Allen Rostron & Nancy Levit, Information for Submitting Articles to Law Reviews & Journals (July 2022) (unpublished manuscript) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1019029) (listing law reviews with an anonymous screening policy); Lindgren, *supra* note 4, at 538 (proposing that law reviews "[c]onceal the author's identity, gender, and institutional affiliation from those selecting the articles").

including limiting the practice of simultaneous submission or adopting peer review, have not gained traction in law review submissions despite their prevalence in fields like science, technology, and medicine.[9]

This Article is the first to describe how editors might use artificial intelligence technology in the law review submissions process, discussing its promise and pitfalls. Artificial intelligence has been defined as "a set of techniques aimed at approximating some aspect of human or animal cognition using machines."[10] It cannot engage in the logical and critical reasoning skills necessary for substantively evaluating submissions.[11] Despite its deficiencies relative to human intelligence, artificial intelligence can partially automate at least some portion of the tasks associated with the selection process. It can simplify the process of checking for preemption and detecting plagiarism.[12] Editors can use screening tools to ensure that a submission respects formatting requirements and that an author did not fail to cite relevant articles.[13] A partially automated system could assign a score to indicate

[9] *See* Friedman, *supra* note 1, at 1352 (proposing mechanisms to improve the article selection process, including requiring acceptance of the first offer received and limiting the number of simultaneous submissions); Tietz & Price, *supra* note 4, at 314-15 (noting that the practice of "formal peer review" generally does not exist in legal scholarship); Wise et al., *supra* note 8, at 73 (describing a "Peer Reviewed Scholarship Marketplace" that appears to be no longer in operation).

[10] Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 UC DAVIS L. REV. 399, 404 (2017).

[11] The viability of such technology is questionable, though technology in the artificial intelligence area is rapidly evolving. *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 88-89 (2014) [hereinafter *Machine Learning*] ("In the last few decades, researchers have successfully used machine learning to automate a variety of sophisticated tasks that were previously presumed to require human cognition."); Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135, 1182-91 (2019) (describing the potential effects of artificial intelligence technology on judging); GPT-3, *A Robot Wrote This Entire Article. Are You Scared Yet, Human?*, GUARDIAN (Sept. 8, 2020, 4:45 EDT), https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3 [https://perma.cc/WJ58-YG54] (publishing an article written by OpenAI's language generator, GPT-3); *Service*, QATENT, https://qatent.com/service (last visited Dec. 30, 2021) [https://perma.cc/9HP5-NVC7] (stating that its product can "in one click, generate a description ready to be filed" based on a draft of patent claims).

[12] *Home*, SCHOLARSIFT, https://www.scholarsift.com (last visited Dec. 21, 2021) [https://perma.cc/3X5D-ZZTT]; *Plagiarism Checker by Grammarly*, GRAMMARLY, https://www.grammarly.com/plagiarism-checker (last visited Jan. 29, 2022) [https://perma.cc/3SCC-DRA3] ("Grammarly's plagiarism checker detects plagiarism . . . ."); *Turnitin AI*, TURNITIN, https://www.turnitin.com/ai (last visited Jan. 29, 2022) [https://perma.cc/869W-UYHF] ("Using AI to compare writing styles, Turnitin insights help indicate authorship.").

[13] *Home*, *supra* note 12; *see Online Citation Generators*, UNIV. OF WASH. GALLAGHER L. LIBR., https://liblawuw.libguides.com/c.php?g=1236949&p=9051803 (last visited

whether submissions meet certain requirements, such as word count, increasing the efficiency of initial screening. Although a law review editor could examine submissions in a similar manner, technology-assisted review would do so in a more efficient way and allow for evaluation of a greater number of variables.[14] By decreasing the time spent triaging submissions, a partially automated system may give editors additional time to focus their energy on those submissions that are likely to benefit from further evaluation and to engage in anonymized review on a broader scale.

In addition to assessing compliance with superficial requirements, a partially automated system could help editors assess the completeness of research in a given submission.[15] Artificial intelligence can detect whether a submission overlooks seminal cases, statutes, or articles in specific areas of law.[16] Technology-assisted review could augment the ability of human editors, constrained by a narrower knowledge base, to recognize the depth of research that a given submission encapsulates.[17]

Artificial intelligence can also provide a mechanism to evaluate the current submissions process, potentially uncovering bias in human editors.[18] For instance, information about the author's institutional affiliation is unlikely to indicate the quality of a submission on an individual basis, and its use may result in discriminatory outcomes based on race or gender.[19] If an editor otherwise would have declined

Feb. 2, 2022) [https://perma.cc/RB6T-GED6]; *Text Analyzer*, JSTOR LABS, https://www.jstor.org/analyze (last visited Apr. 21, 2022) [https://perma.cc/5FEH-HTJ6].

[14] *Cf.* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 UC DAVIS L. REV. 653, 670-71 (2017) (describing how machine learning considers correlations between variables in making predictions).

[15] *Cf. Home*, *supra* note 12 (describing the use of artificial intelligence to assess "strengths and weaknesses in your research").

[16] *Id.*; *Text Analyzer: About*, JSTOR LABS, https://www.jstor.org/analyze/about (last visited Apr. 21, 2022) [https://perma.cc/Y6Y4-HGXX] ("[Text Analyzer] analyzes the text within the [uploaded] document to find key topics and terms used, and then uses the ones it deems most important — the 'prioritized terms' — to find similar content in JSTOR.").

[17] *See infra* Part II.A.1.

[18] *See* Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 634 (2017) (describing how it is possible to "'peer into the brain' of an algorithm" to ascertain bias); *cf.* Dominik Hangartner, Daniel Kopp & Michael Siegenthaler, *Monitoring Hiring Discrimination Through Online Recruitment Platforms*, 589 NATURE 572 (2021) (discussing a study that used machine learning to monitor discrimination by recruiters using employment websites).

[19] *See infra* Part II.A.2.

to engage in further review of a submission based on questionable heuristics, a high score assigned by an algorithm might indicate that implicit bias could be affecting the editor's selection decision.[20] Revealing human biases may enable development of ways to address their effects.[21] A less biased system of selecting articles for publication could improve the submissions experience for both editors and authors, promoting inclusiveness and enhancing the quality of legal scholarship.[22]

Screening software is already being used to manage the torrent of documents in numerous fields. During discovery in litigation, attorneys use technology-assisted review to filter potentially millions of documents that are unlikely to be relevant or to identify a smaller subset of documents that need further review to determine if privilege exists.[23] Employers regularly use hiring algorithms to assess whether job

[20] *See* Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Algorithms as Discrimination Detectors*, 117 PROC. NAT'L ACAD. SCI. 30096, 30100 (2020) ("Introducing a data-driven decision aid creates additional opportunities to detect what the humans in the system are doing, since we can test whether human compliance with the tool's recommendations, as opposed to override, is systematically lower or higher for protected groups.").

[21] *See* Kroll et al., *supra* note 18, at 634.

[22] *Cf.* Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 572 (2018) (concluding that "technology has the potential to improve upon human decision-making by suppressing or removing human biases"); Alex P. Miller, *Want Less-Biased Decisions? Use Algorithms*, HARV. BUS. REV. (July 26, 2018), https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms [https://perma.cc/7K3D-RQRR] ("Algorithms deliver more-efficient *and* more-equitable outcomes."). *But cf.* Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1696-99 (2020) (describing the problems of relying on automated decision-making as a way to address human bias).

[23] *See* Herbert L. Roitblat, Anne Kershaw & Patrick Oot, *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 J. AM. SOC'Y INFO. SCI. & TECH. 70, 79 (2010) (concluding that categorizing relevant or responsive documents by computer systems is "at least as accurate" as using human reviewers); Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1329-30 (2019) [hereinafter *Artificial Intelligence and the Law*] ("This automated-review software became necessary with the rise of e-discovery, as the document troves related to particular lawsuits began to rise into the hundreds of thousands and sometimes millions of documents . . . ."); Maggie Burtoft, *Electronic Discovery Document Review: The Power of Feedback*, LAW.COM (Oct. 21, 2021, 11:50 AM), https://www.law.com/2021/10/21/electronic-discovery-document-review-the-power-of-feedback [https://perma.cc/H9D6-HDPS] (explaining that "discovery teams can curb review costs by employing early case assessment methods to reduce the volume of documents requiring review and technology-assisted review to identify responsive documents").

applications should receive further consideration.[24] More closely related, many scientific journals have incorporated artificial intelligence into their submissions screening processes.[25] The use of such techniques has drawn criticism, however, because of their potential to replicate existing biases.[26] For example, a hiring algorithm learned to penalize job applications that mentioned the word "women's" because developers trained it using resumes primarily submitted by men.[27]

Technology-assisted review could similarly complement in-depth substantive review by law review editors,[28] but its use raises significant concerns. A partially automated system may negatively impact certain groups that are not adequately represented in the process of training the system.[29] For example, if developers use examples of publications from a limited number of fields of study, the system will not be able to accurately assign scores for research comprehensiveness to submissions in less common topic areas. Incompleteness in the representativeness of data is amplified because predictive technology relies on a limited amount of data to draw broad conclusions.[30] To the extent developers

[24] *See* Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARV. BUS. REV. (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias [https://perma.cc/EMF9-X67N] (describing the different applications of algorithms in the hiring process).

[25] See *infra* Part I.B.1.

[26] *See* Ajunwa, *supra* note 22, at 1708 (explaining that "the biased results of algorithmic hiring systems . . . reveal legal anachronisms, such as an American tradition of deference to the employer"); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 (2016) ("Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society."); Bogen, *supra* note 24 (describing how hiring algorithms can "replicate institutional and historical biases"); *infra* Parts I.B.1, II.B.

[27] *See* Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, REUTERS (Oct. 10, 2018, 4:04 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/2H7G-6DY7]; *infra* Part II.B.2.

[28] *See* Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 46-48 (2019) (discussing how artificial intelligence could assist human intelligence without displacing it); Volokh, *supra* note 11, at 1149 (describing how "early AIs will be aimed at helping human decisionmakers . . . rather than supplanting them").

[29] *See* Barocas & Selbst, *supra* note 26, at 677-93 (describing how discriminatory outcomes may occur when a machine learning system is trained using data that is not representative); *infra* Part II.B.2.

[30] *See* Barocas & Selbst, *supra* note 26, at 686-90.

train the system based on past determinations of article strength, such as only highly-cited articles or those published in top ten law reviews, the system may not accurately assign a score for submissions in emerging areas of law or those that use unique writing styles.[31] The opacity of many types of partially automated systems, coupled with misplaced deference to seemingly neutral algorithms, magnifies the risk of adverse outcomes.[32]

Individuals involved in developing the technology may have biases, which they unintentionally incorporate into the technology.[33] The system might assign lower scores to papers based on features that contain embedded bias or proxies for protected characteristics of authors, such as the law school the author attended.[34] Developers might also use data that contains systemic bias to teach the technology how to score submissions.[35] For instance, an algorithm that relies on the placement of an author's previous articles in highly-ranked journals to

[31] *See* Lehr & Ohm, *supra* note 14, at 680 (describing the importance of "generalizability — the ability of an algorithm trained on a particular dataset to generate accurate predictions when deployed on different data"); Surden, *Machine Learning*, *supra* note 11, at 105 (describing how a model is "thus only useful to the extent that the heuristics inferred from past cases can be extrapolated to predict novel cases"); *cf.* Gaillynn Clements, *An Unexpected Irony: Lifting the "Diversity" Wool from our Eyes*, *in* LINGUISTIC DISCRIMINATION IN U.S. HIGHER EDUCATION: POWER, PREJUDICE, IMPACTS, AND REMEDIES 1, 4-6 (Gaillynn Clements & Marnie Jo Petray eds., 2021) (describing how "universities set a medium of instruction . . . which is necessary for idea sharing (including publishing)" and that "language discrimination is a proxy for racial, ethnic, regional, social class, and gender discrimination").

[32] *See* Ajunwa, *supra* note 22, at 1686 (explaining how "the belief in data objectivity then often results in an uncritical acceptance of decisions derived from such algorithmic systems"); Douglas Heaven, *AI Peer Reviewers Unleashed to Ease Publishing Grind*, 563 NATURE 609, 610 (2018) (stating that "there might be temptation for editors to cut corners and simply rely on [an assigned] score in deciding to reject a paper"); Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1337 (discussing how decision makers might "inappropriately defer to this false precision, failing to take into account the limits of the model, the uncertainties involved, the subjective decisions that went into the model's creation," and the rate of false negatives); Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 619 (2019) (describing how the "opacity of decision-making algorithms prevents those harmed by automated systems from determining either how a decision came about or the logic and reasoning behind it").

[33] *See* Waldman, *supra* note 32, at 621-22.

[34] *See infra* Part II.A.2.

[35] *See* Barocas & Selbst, *supra* note 26, at 684-87; Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1336 ("There is a concern that automated AI-enhanced decisions may disproportionately appear to be more neutral, objective, and accurate than they actually are.").

predict the strength of a new submission may incorporate racial bias.[36] To the extent that editors currently consider such information as an indication of submission strength, partial automation may simply enable efficient implementation of existing biases rather than correcting them. In addition, a partially automated system may be too costly for widespread adoption, excluding those institutions and editors that might benefit most from it and enabling strategic gaming for authors with institutional access.[37]

With careful development, limited application, and attentive oversight — including conducting regular audits for discriminatory outcomes and requiring measured transparency — some of these harms can be addressed.[38] The potential gains from using artificial intelligence may make an imperfect system worth considering in circumscribed applications.[39] For instance, although one would not want to rely on Google Translate for matters of significance — such as translating a contract — it still provides a resource where one might otherwise not exist, like interpreting a menu at a restaurant.[40] Similarly, editors should not rely on a score assigned by artificial intelligence to decide whether to accept an article. However, technology-assisted review can streamline assessment of stylistic requirements for submissions, allowing editors

---

[36] *See infra* Part II.A.2.

[37] *See infra* Part III.

[38] *See id.*; *cf.* Harini Suresh & John Guttag, *A Framework for Understanding Sources of Harm Throughout the Machine Learning Life Cycle*, ACM DIGIT. LIBR., Oct. 2021, at 7-8, https://dl.acm.org/doi/pdf/10.1145/3465416.3483305 [https://perma.cc/66MR-FJ4R] (setting forth a "framework to provide a useful organizational structure for thinking through potential problems, understanding if and what mitigation techniques are appropriate, and/or motivating new ones").

[39] *See* Surden, *Machine Learning*, *supra* note 11, at 88 ("While the results of these automated efforts are sometimes imperfect, the interesting point is that such computer generated results have often proven useful for particular tasks where strong approximations are acceptable.").

[40] *See id.* at 100 ("Such automation has allowed for approximate but useful translations in many contexts where no translation was previously available at all."); *Translation AI*, GOOGLE CLOUD, https://cloud.google.com/translate/#how-automl-translationbeta-works (last visited July 25, 2021) [https://perma.cc/B2RR-P6PQ]; *cf.* Alessandro Checco, Lorenzo Bracciale, Pierpaolo Loreti, Stephen Pinfield & Giuseppe Bianchi, *AI-Assisted Peer Review*, 8 HUMANS. & SOC. SCIS. COMMC'NS. 1, 2 (2021) (discussing the use of an "Automated Essay Scoring (AES) application" that MIT, Harvard, and EdX use "to assess written work in their MOOCs"); Cade Metz, *Can A.I. Grade Your Next Test?*, N.Y. TIMES (July 20, 2021), https://www.nytimes.com/2021/07/20/technology/ai-education-neural-networks.html [https://perma.cc/S48B-8CZ3] (describing Stanford University's use of automated feedback in one of its online courses).

the time to substantively examine a greater number of submissions.[41] The use of artificial intelligence might also address serious shortcomings in the current selection process, providing feedback to editors about biases in their decision-making and minimizing the overreliance on lower-ranked journals to conduct the initial screening of submissions.[42]

This Article proceeds in three Parts. Part I describes the foundational concepts of using artificial intelligence technology to comb through submissions in a variety of fields. Part II examines the potential benefits and risks of implementing technology-assisted review in the law review submissions process. Part III sets forth possible mechanisms for mitigating the potential harms identified and offers measured suggestions for using artificial intelligence in the law review submissions process. By recognizing the limitations of artificial intelligence, law review editors may be able to use technology-assisted review in ways that provide for a more effective and less biased submissions experience.

I. USING ARTIFICIAL INTELLIGENCE TO HELP SCREEN SUBMISSIONS

Artificial intelligence develops and uses technology to execute tasks that ordinarily depend on human intelligence.[43] Common examples of artificial intelligence include image and speech recognition and language translation.[44] Algorithms are essentially the automated instructions that form the foundation of artificial intelligence.[45] Machine learning is a subset of artificial intelligence that trains machines how to learn to do a variety of things.[46] Machine learning

---

41 *See infra* Part II.C.

42 Joseph Scott Miller, *The Immorality of Requesting Expedited Review*, 21 LEWIS & CLARK L. REV. 211, 214-15 (2017) ("The work [editors at lower ranked journals] do is redistributed up the prestige hierarchy, never to return.").

43 IBM Cloud Education, *Artificial Intelligence (AI)*, IBM (June 3, 2020), https://www.ibm.com/cloud/learn/what-is-artificial-intelligence [https://perma.cc/8V9N-JHHZ].

44 *Id.*; *see* Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1307 ("Researchers have successfully applied AI technology to automate some complex activities, including playing chess, translating languages, and driving vehicles.").

45 Stephen F. DeAngelis, *Artificial Intelligence: How Algorithms Make Systems Smart*, WIRED (Sept. 2014), https://www.wired.com/insights/2014/09/artificial-intelligence-algorithms-2 [https://perma.cc/M677-UADH].

46 MEHRYAR MOHRI, AFSHIN ROSTAMIZADEH & AMEET TALWALKAR, FOUNDATIONS OF MACHINE LEARNING 1-3 (2d ed. 2018); *Machine Learning: What It Is and Why It Matters*, SAS, https://www.sas.com/en_us/insights/analytics/machine-learning.html (last visited July 15, 2021) [https://perma.cc/U3DP-SL6S].

methods work by detecting patterns in vast amounts of data and then applying the patterns to accomplish a given task — from recommending videos to driving a vehicle.[47] For example, a consumer's credit card company analyzes data about purchases using machine learning.[48] Information related to the date, location, and amount of each transaction is structured data that helps the machine learning algorithm define a typical transaction for a given user. Consequently, if a user typically makes purchases in New York, the machine learning algorithm will view purchases in Texas as a deviation, and the credit card company may stop the transaction.

Artificial intelligence systems sometimes appear to complete tasks that involve cognitive abilities.[49] They perform these tasks using proxies to represent foundational elements, concepts, or features.[50] For example, machine learning can use pattern detection to infer proxies for lower quality emails, such as the use of all capital letters or country of origin to detect spam.[51] Thus, even though the technology does not examine the meaning of the email itself, it can still be useful in filtering email.[52] It is difficult for these systems, however, to classify text associated with an abstract objective.[53]

Although algorithms are mathematical, they are not neutral. In a supervised learning approach, the data used to train the machine learning algorithms is carefully selected, organized, and validated by

[47] *Machine Learning: What It Is and Why It Matters*, *supra* note 46; *see* Sharona Hoffman & Andy Podgurski, *Artificial Intelligence and Discrimination in Health Care*, 19 YALE J. HEALTH POL'Y, L., & ETHICS 1, 8 (2020).

[48] *Machine Learning: What It Is and Why It Matters*, *supra* note 46.

[49] *See* Volokh, *supra* note 11, at 1137-38 (arguing that the focus should not be "whether we recognize its reasoning processes as intelligent," but "whether the output of those processes provides what we need"); Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1330 (describing how artificial intelligence is not able to make decisions that "involve understanding the law and the facts and dealing with strategy, policy, and other abstractions"). *But see* Brian Haney, *Patents for NLP Software: An Empirical Review*, 18 IUP J. KNOWLEDGE MGMT. 27, 37 (2020) (explaining that "developing software exhibiting common sense reasoning at human level is a formidable problem").

[50] *See* Surden, *Machine Learning*, *supra* note 11, at 95.

[51] *See id.* at 96-97.

[52] *See id.*

[53] *See id.* at 113 (explaining that "algorithms are not well suited to, or intended to, apply legal judgment in nuanced, uncertain areas"); Haney, *supra* note 49, at 40 ("NLP software processing information often assists human decision makers, rather than making decisions autonomously.").

people.[54] Unlike traditional algorithms, in which a developer explicitly programs the decision-making rules, developers in machine learning systems select the training data (past examples) and label outputs.[55] The training data then enables the creation of a model that the machine learning system can use to analyze new situations and make predictions.[56] Machine learning algorithms are able to create more intricate models by discovering new patterns in the data.[57] As the machine learning algorithms evaluate more data and identify additional patterns in the data, their decision-making ability improves.[58] To be effective, machine learning requires "large amounts of high-quality, structured, machine-processable data," so it will not perform as well where data is lacking in quantity or quality.[59] The better the data, the better the algorithm should perform — absent any bias that has made its way into the learning process.[60]

Developers use natural language processing ("NLP") techniques in language-focused artificial intelligence systems; they too require enormous quantities of human-provided data to be effective.[61]

---

54 This Article will focus on supervised learning approaches. *See* Surden, *Machine Learning*, *supra* note 11, at 91-94 (offering a spam filtering algorithm as an example of a supervised learning approach); *cf.* Barocas & Selbst, *supra* note 26, at 678 n.24 ("Other techniques known as 'unsupervised' learning do not require any such target variables and instead search for general structures in the dataset, rather than patterns specifically related to some state or outcome."); Lehr & Ohm, *supra* note 14, at 676 ("Unsupervised learning algorithms do not predict outcome variables labeled with ground truth. Instead, they group or cluster subjects together based, roughly speaking, on how similar their input data values are.").

55 *See* Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEO. L.J. 803, 809 (2020); Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1395 (2019).

56 *See* Surden, *Machine Learning*, *supra* note 11, at 91-94.

57 *See id.* at 94; Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1279 (2018) (explaining that machine learning allows for the "extraction of implicit knowledge by discovering patterns or relationships within a data set").

58 *See* Surden, *Machine Learning*, *supra* note 11, at 93-94 ("Such algorithms are powerful because, in a sense, these algorithms program themselves over time with the rules to accomplish a task, rather than being programmed manually with a series of pre-determined rules."); *What Is Machine Learning?: How It Works, Why It Matters, and Getting Started*, MATHWORKS, https://www.mathworks.com/discovery/machine-learning.html (last visited July 15, 2021) [https://perma.cc/VVP3-HJYZ].

59 Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1316.

60 *See id.* at 1316, 1335-36.

61 IBM Cloud Education, *Natural Language Processing (NLP)*, IBM (July 2, 2020), https://www.ibm.com/cloud/learn/natural-language-processing [https://perma.cc/32AT-9XYV]; *see also* Prakash M. Nadkarni, Lucila Ohno-Machado & Wendy W. Chapman, *Natural Language Processing: An Introduction*, 18 J. AM. MED. INFO. ASS'N 544, 544-45 (2011) (providing an overview of NLP and its limitations).

Developers can train an NLP model to examine individual words, the sequence and grouping of words, as well as the layout and formatting of a document.[62] The NLP model could then process a new document.[63] Systems using NLP techniques may face difficulty evaluating complicated documents.[64]

The following discussion will touch on the use of artificial intelligence in screening submissions in a variety of areas, and then describe its current application in the submissions process for scientific journals and law reviews.

A. Examples of Screening Technology in Other Fields

Artificial intelligence technology can be trained to sort through vast quantities of data in an efficient and effective way. This Section provides examples of how automated technology can help sift through documents during litigation discovery, assist in examination of patent applications by the United States Patent and Trademark Office ("Patent Office"), and filter resumes for further review in the hiring process. Although these partially automated mechanisms increase efficiency, they also raise serious concerns about discrimination and bias.[65]

In civil litigation, parties often exchange volumes of documents during the discovery process.[66] Discovery is the means of gathering and reviewing evidence in litigation.[67] Prior to the emergence of electronic discovery ("e-discovery"), armies of junior attorneys used to sift through troves of materials to determine if a document was potentially relevant to the matter at issue or privileged or both.[68] Electronic discovery facilitated the development of partially automated document

---

[62] *See* Spencer Williams, *Predictive Contracting*, 2019 COLUM. BUS. L. REV. 621, 653.

[63] *See id.*

[64] Checco et al., *supra* note 40, at 4.

[65] *See infra* Part II.B.

[66] *See* Burtoft, *supra* note 23; Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1329-30.

[67] Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1329-30.

[68] *See id.* (noting that "human review of documents will continue to play a huge part in the e-discovery process"); MARTIN FORD, RISE OF THE ROBOTS: TECHNOLOGY AND THE THREAT OF A JOBLESS FUTURE 124 (2015) (describing the replacement of lawyers and paralegals by "powerful algorithms that can analyze millions of documents and automatically tease out" those that are likely to be relevant); Charles Yablon & Nick Landsman-Roos, *Predictive Coding: Emerging Questions and Concerns*, 64 S.C. L. REV. 633, 637 (2013) ("The use of technology-assisted review began around 2008, when a small number of law firms started exploring ways in which they could use computers and sophisticated software to make the discovery review process more efficient.").

review, known as "predictive coding" or "technology-assisted review."[69] Predictive-coding systems are able to filter massive amounts of discovery documents — sometimes numbering in the millions — to a manageable amount of material by removing those that are very likely to be irrelevant.[70] These systems are not very useful for determining if a document is likely or even slightly relevant, but they are helpful in figuring out the documents that are very likely not relevant.[71] Attorneys then review the subset of documents highlighted as needing further review to determine if the documents are in fact relevant or qualify for an assertion of privilege.[72] By using patterns to sort documents, the predictive-coding technology is able to remove the documents that are least likely to be relevant to the matter at issue, preserving valuable time for attorney review of only those documents that are more likely to be relevant.[73]

The use of predictive coding in e-discovery exposes some limitations in using technology-assisted review to help classify documents. Although predictive coding technology can identify discovery documents likely to be irrelevant, the ultimate decision of relevance or privilege must be made by human attorneys.[74] Within e-discovery, there are some tasks that lend themselves to partial automation.[75] There may be certain rules that an artificial intelligence system can recognize. For example, in a matter involving employment discrimination, the reviewing attorney could train the predictive coding technology to search for terms that might raise concern, such as expletives, or the technology might learn new terms based on its analysis of previous discrimination cases.[76]

---

69 Yablon & Landsman-Roos, *supra* note 68, at 637.

70 *Id.*

71 *See* Surden, *Machine Learning*, *supra* note 11, at 113.

72 *Id.* (explaining that "the algorithms perform the role of filtering down the size of the document stack that is ultimately in need of lawyerly review").

73 Yablon & Landsman-Roos, *supra* note 68, at 638.

74 Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1330.

75 *See, e.g.*, Surden, *Machine Learning*, *supra* note 11, at 104 (describing how "an algorithm may identify a complex mix of factors in the data associated with particular outcomes that may be hard or impossible for an attorney to detect using typical legal analysis methods"); *About Us*, LEX MACHINA, https://lexmachina.com/about (last visited Jan. 2, 2022) [https://perma.cc/7E92-VPQ7] (describing legal analytics software, which "mines and processes litigation data, revealing insights never before available about . . . the subjects of the cases themselves, culled from millions of pages of litigation information").

76 *See* Surden, *Machine Learning*, *supra* note 11, at 105; *About Us*, *supra* note 75.

Similarly, the Patent Office uses artificial intelligence to help manage a massive volume of applications filed each year.[77] Patent examiners use machine learning technology to enhance their ability to identify documents relevant to evaluating applications.[78] In this context, machine learning algorithms can organize the immense amount of prior art. Prior art is defined as the materials that set forth the state of the art at the time the patent application was filed.[79] Searching for and reviewing prior art is a time-consuming yet essential undertaking, as it is critical in assessing whether an invention is sufficiently new and nonobvious to deserve patent protection.[80] Although the Patent Office uses machine learning to limit the expanse of prior art, patent examiners must ultimately decide whether a patent should be granted after evaluating the prior art and comparing it to the application at issue.[81]

As another example, computerized sorting mechanisms are widely used to narrow the pool of applications in the hiring process.[82] Algorithms identify candidates that companies might want to pursue, and they are used to target job advertisements to a select group of potential candidates.[83] Predictive technologies can help employers sift

---

[77] Arti K. Rai, *Machine Learning at the Patent Office: Lessons for Patents and Administrative Law*, 104 IOWA L. REV. 2617, 2619 (2019) ("The Patent Office receives hundreds of thousands of patent applications every year, and the examiners who process the applications operate under severe time pressure.").

[78] *See, e.g.*, Lehr & Ohm, *supra* note 14, at 671 (defining machine learning as "an automated process of discovering correlations (sometimes alternatively referred to as relationships or patterns) between variables in a dataset, often to make predictions or estimates of some outcome"); Rai, *supra* note 77, at 2618 (describing the use of machine learning at the Patent Office); Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1311 (defining machine learning as "a family of AI techniques" that "work by detecting useful patterns in large amounts of data").

[79] *See* Rai, *supra* note 77, at 2619, 2634 ("[M]achine learning could be particularly useful for the time-intensive but critical task of searching the prior learning ('prior art') to determine whether, at the time of patent filing, the invention claimed was novel and nonobvious."); Brenda M. Simon, *Rules, Standards, and the Reality of Obviousness*, 65 CASE W. RSRV. L. REV. 25, 28 (2014) (describing how "increased access to searchable information and processing power provides additional time to consider a wider range of prior art").

[80] *See* 35 U.S.C. §§ 102(a), 103 (2018); Rai, *supra* note 77, at 2619, 2634 (describing the use of machine learning in prior art searching); Brenda M. Simon, *The Implications of Technological Advancement for Obviousness*, 19 MICH. TELECOMM. & TECH. L. REV. 331, 346 (2013) (explaining that "the quality of prior art located for a given application is limited by time, ability, interest, and resources").

[81] *See* Rai, *supra* note 77, at 2631.

[82] Bogen, *supra* note 24.

[83] *Id.*

through a deluge of applications. For better or worse, almost three-quarters of resumes never make it past the digital gatekeepers.[84] Companies score, rank, and screen applicants using application tracking systems.[85] Software compares applicants against the profile of an "ideal candidate."[86] To increase scores, commentators have suggested candidates use standard fonts, remove images, and include standard industry acronyms as well as language found in the job listing.[87] Researchers have described how the use of hiring algorithms reproduces existing biases and has had discriminatory effects, which will be described in greater detail below.[88]

B. Technology-Assisted Review of Journal Submissions

Editors in different fields already use some form of partially automated technology in reviewing submissions. This Section details the current use of artificial intelligence by scientific journals and law reviews. Even partial automation of certain aspects of the submissions process carries the potential for substantial harm.

1. Using Artificial Intelligence in Scientific Journal Submissions

The volume of submissions has strained the feasibility of peer review for scientific journals.[89] Editors have already started to adopt partially automated screening tools, though their use has raised significant concerns.[90] Modest advances in technology have enabled journals to reduce screening time by confirming a submission complies with

---

[84] Mona Abdel-Halim, *12 Ways to Optimize Your Resume for Applicant Tracking Systems*, MASHABLE (May 27, 2012), https://mashable.com/2012/05/27/resume-tracking-systems [https://perma.cc/FE78-LCPV]; *see infra* Part II.B.3.

[85] Abdel-Halim, *supra* note 84.

[86] Jessica Leber, *The Machine-Readable Workforce*, MIT TECH. REV. (May 27, 2013), https://www.technologyreview.com/2013/05/27/178320/the-machine-readable-workforce [https://perma.cc/Z6U6-BQYE].

[87] Abdel-Halim, *supra* note 84.

[88] *See* Dastin, *supra* note 27; Cathy O'Neil, *Amazon's Gender-Biased Algorithm Is Not Alone*, BLOOMBERG (Oct. 16, 2018, 6:00 AM PDT), https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone [https://perma.cc/72T4-78BE]; *infra* Part II.B.2.

[89] Checco et al., *supra* note 40, at 2 (stating that submissions have grown annually by 6.1% since 2013, and estimating that "over 15 million hours are spent every year on reviewing of manuscripts previously rejected and then resubmitted to other journals").

[90] *Id.*; Jef Akst, *Researchers to CIHR: Reverse Peer Review Changes*, SCIENTIST (July 5, 2016), https://www.the-scientist.com/the-nutshell/researchers-to-cihr-reverse-peer-review-changes-33236 [https://perma.cc/HS5A-MEYD].

formatting requirements.[91] More sophisticated software, such as statcheck or StatReviewer, can verify statistical analysis and methods, although some have questioned their validity and reliability.[92] Editors use other technologies to locate suitable peer reviewers or serve as an adjunct in quality screening.[93]

One of the more sophisticated automated tools, UNSILO, claims to be able to analyze the content of submissions using machine learning and natural language processing; it attempts to glean the main points from a submission to create a summary.[94] Rather than using the authors' self-identified key words, UNSILO attempts to extract key concepts to summarize results, ascertain the likelihood of plagiarism, and determine how the submission relates to the broader literature.[95] These systems must examine a large corpus of documents to analyze new

---

[91] *See* Checco et al., *supra* note 40, at 2-3; *see, e.g.*, CLARIVATE, https://clarivate.com/webofsciencegroup/solutions/scholarone (last visited Dec. 21, 2021) [https://perma.cc/9RHM-6K22] (discussing "AI-powered metadata extraction and submission filtering"); OVERLEAF, https://www.overleaf.com/for/publishers#benefits (last visited Dec. 21, 2021) [https://perma.cc/7MCX-AFDB] (stating that the "automated Overleaf pre-submission check system" ensures submissions are correctly formatted); PENELOPE, https://www.penelope.ai (last visited Dec. 21, 2021) [https://perma.cc/SDL9-QDRM] (describing a resource that can assess whether the citations and format of a submission comply with the journal's requirements).

[92] STATCHECK, http://statcheck.io (last visited Dec. 21, 2021) [https://perma.cc/W22B-ZEW2]; STATREVIEWER, http://www.statreviewer.com (last visited Dec. 21, 2021) [https://perma.cc/4DWL-6ENP]; *see also* Heaven, *supra* note 32, at 609-10 (describing how statcheck "assesses the consistency of authors' statistics reporting" while StatReviewer "checks that papers correctly include things such as sample sizes, information about blinding of subjects and baseline data"); Thomas Schmidt, *Statcheck Does Not Work: All the Numbers. Reply to Nuijten et al.*, PSYARXIV (Nov. 22, 2017), https://psyarxiv.com/hr6qy [https://perma.cc/HZZ5-UXYR] (raising concerns about the validity and reliability of statcheck).

[93] *See, e.g.*, Checco et al., *supra* note 40, at 2 (describing the use of an "Automated Essay Scoring (AES) application, used by EdX, MIT and Harvard's non-profit MOOC federation to assess written work in their MOOCs"); Lorcan Reilly, *About Web of Science Reviewer Locator*, CLARIVATE, https://publons.freshdesk.com/support/solutions/articles/12000047301-about-web-of-science-reviewer-locator (last updated Aug. 18, 2022) [https://perma.cc/QDX5-337V] (describing its services as a way to "find, screen, and connect with the subject matter experts needed to peer review manuscript submissions").

[94] *Technology*, UNSILO, https://web.archive.org/web/20211127073000/https://unsilo.ai/technology/ [https://perma.cc/5ZNR-N9QQ] (last visited Aug. 16, 2022); Heaven, *supra* note 32, at 609.

[95] *Technology*, *supra* note 94; Heaven, *supra* note 32, at 609 (describing the development of "software that can mine paper databases and extract connections between different disciplines and concepts"); *see also* WIZDOM.AI, https://www.wizdom.ai (last visited Dec. 21, 2021) [https://perma.cc/WKV5-Y3EF].

papers adequately.[96] In comparing submissions with previously published works, UNSILO draws upon 1.7 million research papers from the PubMed Central database, and claims that it is in the process of augmenting its databases with an additional twenty million papers.[97]

As an example of the possibilities and risks of using artificial intelligence in screening submissions, a recent empirical study examined whether an artificial intelligence tool could successfully predict the peer review score of an unreviewed manuscript through the use of its textual content alone.[98] A group of researchers designed and trained the tool using 3,300 papers and the peer review evaluations of those papers.[99] They trained it to evaluate papers based on formatting, textual content (the frequency with which words were used), and readability (the size of words, complexity of vocabulary, and sentence length).[100] Despite such a superficial review, the system was frequently able to predict the peer review results.[101] The authors found the existence of correlations between decision-making and the limited quality proxy measures described above, suggesting that partial automation may be able to play a role in the review process.[102] They also describe how a partially automated system would allow reviewers "to focus more on the scientific content" of a given submission.[103]

The ability to predict peer review scores based on a superficial examination of an unreviewed manuscript may indicate bias in the review process.[104] A system using artificial intelligence to screen papers might have a higher rejection rate than a standard peer review system for papers on "innovative topics" or those that contain characteristics associated with lower-income countries.[105] Papers from disadvantaged regions have been historically underrepresented in the literature, and the system might not recognize the improving quality of papers submitted from these regions over time.[106] If the system is trained on papers selected by editors who have relied on American reviewers that

---

96 *Technology*, *supra* note 94; *see* Heaven, *supra* note 32, at 609.
97 *Technology*, *supra* note 94.
98 Checco et al., *supra* note 40, at 3.
99 *Id.* at 1.
100 *Id.* at 3.
101 *Id.*
102 *Id.*
103 *Id.* at 9.
104 *Id.*
105 *Id.*
106 *Id.*

tend to favor papers from higher-income regions,[107] an automated system may perpetuate biases against submissions from lower-income countries.[108] Other researchers have also described how screening technology that has been trained using published papers might perpetuate biases that already exist in peer review.[109]

Many in the scientific research community have raised other concerns about the potential for adverse outcomes in using automated techniques to evaluate submissions.[110] They have criticized the validity and reliability of automated technology used to assess the consistency of a submission's statistical analysis.[111] Researchers also worry that if the automated screening technology allocates a single score after evaluation, editors could place too much weight on the score in rejecting submissions.[112] In the context of grant review, almost 1,200 researchers criticized the decision of the Canadian Institutes of Health Research to transition from in-person meetings for peer review panels to an online system for grant evaluation.[113] Researchers found that in the new online evaluation, female applicants "fared significantly worse than their male counterparts, and younger researchers received less money than more-senior faculty" as compared with the prior system.[114] They believe the quality of the review process suffered, at least in part, from the failure of reviewers to evaluate the proposals in consultation with the other scientists.[115]

---

[107] *Id.* (describing how the United States "dominates the contribution to peer review").

[108] *See id.* (explaining that "a model may propagate cultural and organisational biases already present in the learning set").

[109] Heaven, *supra* note 32, at 610.

[110] *See, e.g.*, *id.* (discussing the "potential pitfalls to AI in peer review in general"); Cassidy R. Sugimoto, *Scientific Success by Numbers*, 593 NATURE 30, 31 (2021) (criticizing the use of journal impact factors); Schmidt, *supra* note 92, at 4 ("Statcheck has low validity, misses many inconsistent tests and makes many false alarms.").

[111] *See* Schmidt, *supra* note 92, at 4.

[112] Heaven, *supra* note 32, at 610 (describing the temptation for editors to "cut corners" and rely on assigned scores in rejecting papers); *cf.* Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/C8FA-3ZM4] (stating that although "judges are not supposed to give longer sentences to defendants with higher risk scores . . . [they] have cited scores in their sentencing decisions").

[113] Akst, *supra* note 90.

[114] *Id.*

[115] *Id.*

More than 20,000 individuals and institutions have signed the Declaration on Research Assessment ("DORA").[116] In DORA, they argue for the research community to stop depending on indicators that may incorporate structural biases,[117] such as racism and sexism. They seek to improve the approaches for assessing research outcomes and emphasize the importance of evaluating research on its own merits, as opposed to where the research was published.[118] The use of technology-assisted review in the context of law review submissions, which will be discussed below, is likely to raise similar concerns.

2. The Current Use of Artificial Intelligence in Law Review Submissions

In the context of law review submissions, although electronic submissions platforms have long been the standard,[119] the use of artificial intelligence is still in its infancy. The primary submissions service, Scholastica, provides technology that "makes reviewing incoming papers a lot faster."[120] It enables editors to examine and filter submissions by author, title, date, keywords, or tags.[121] Editors have many options for services that use artificial intelligence technology in detecting plagiarism and correcting grammar.[122] In addition, they can use tools to confirm that authors have appropriately formatted their citations.[123]

Another company, ScholarSift, states it is the "first analytical platform to deploy machine learning to legal scholarship."[124] Similar to the Patent Office's use of machine learning to identify prior art related to a given patent application, ScholarSift offers technology to search for relevant literature, as well as assess the possible preemption of a submission.[125] Editors and authors can use ScholarSift to confirm that the author did

---

[116] *Signers*, THE DECLARATION ON RSCH. ASSESSMENT (DORA), https://sfdora.org/signers (last visited Feb. 4, 2022) [https://perma.cc/5DGA-CC5C].

[117] *The Declaration*, THE DECLARATION ON RSCH. ASSESSMENT (DORA), https://sfdora.org/read (last visited Feb. 4, 2022) [https://perma.cc/VTC5-Y3YC].

[118] *Id.*

[119] *See, e.g.*, Friedman, *supra* note 1, at 1301 ("Electronic submission services . . . have caused submissions to skyrocket.").

[120] *Law Review System*, *supra* note 2.

[121] *Id.*

[122] *See Plagiarism Checker by Grammarly*, *supra* note 12; *Turnitin AI*, *supra* note 12.

[123] *See Online Citation Generators*, *supra* note 13.

[124] *About Us*, SCHOLARSIFT, https://www.scholarsift.com/about (last visited Dec. 21, 2021) [https://perma.cc/YZ6L-PAM3].

[125] *Home*, *supra* note 12.

not overlook articles relevant to a given submission.[126] The technology compares the text and footnotes in an uploaded paper to a large database of published law journal articles to evaluate similarities, identifying related scholarship and citations that may be pertinent.[127] In future development, ScholarSift claims to be designing technology to filter thousands of submissions to locate the "most promising" papers.[128] It has not disclosed, however, the criteria that its proprietary system will use in determining whether a given submission is "promising."[129] The filtering mechanism does not appear to be available for purchase at the present time.

Relatedly, JSTOR is a non-profit organization that offers a teaching and research platform.[130] It is in the process of developing a somewhat similar technology to the ScholarSift product called the Text Analyzer.[131] As with ScholarSift, users upload a draft, and the tool identifies related materials.[132] Text Analyzer also suggests keywords that authors can use in posting their papers to an online repository.[133] Analogous services exist for practitioners and judges to analyze briefs and other legal materials.[134]

---

126 *See Robert Anderson on Analytics for Law Review Submissions and Publishing*, IPSE DIXIT, at 5:50-6:23 (Feb. 16, 2021), https://shows.acast.com/ipse-dixit/episodes/robert-anderson-on-analytics-for-law-review-submissions-and- [https://perma.cc/UJR2-FDJF] (stating that ScholarSift can make sure that authors can "find relevant literature" for an uploaded paper and "don't miss anything").

127 *Home*, *supra* note 12; *see also* Bonnie Shucha, *Representing Law Faculty Scholarly Impact: Strategies for Improving Citation Metrics Accuracy and Promoting Scholarly Visibility*, 40 LEGAL REFERENCE SERVS. Q. 81, 103 (2021) ("At present, ScholarSift primarily contains freely available articles drawn from Law Review Commons and supplemented by journal websites."); Brian Frye, *It's the End of Citation as We Know It & I Feel Fine*, TECHDIRT (Mar. 22, 2021, 3:49 PM), https://www.techdirt.com/articles/20210318/22393446451/end-citation-as-we-know-it-i-feel-fine.shtml [https://perma.cc/VJ7R-DZT4] (explaining that "ScholarSift tells authors which articles they should be citing").

128 *Home*, *supra* note 12.

129 Email from ScholarSift to author (Mar. 5, 2021, 12:09 PM) (on file with author) (recognizing that ScholarSift will "need to explain aspects of how it works in order to gain confidence from users"); *Robert Anderson on Analytics for Law Review Submissions and Publishing*, *supra* note 126.

130 *About JSTOR,* JSTOR, https://about.jstor.org (last visited Oct. 3, 2022) [https://perma.cc/LDY9-V93A].

131 *See Text Analyzer: About*, *supra* note 16.

132 *Id.*

133 *Id.*; *see also* Shucha, *supra* note 127, at 107.

134 *See, e.g.*, CASETEXT, https://casetext.com/cara-ai (last visited May 11, 2022) [https://perma.cc/PG9X-MB7Y] ("Upload a document from your case to CARA A.I. to find

In summary, editors can use the technological tools described above for identifying failure to comply with formatting requirements, detecting plagiarism, assessing readability, ascertaining missing references, and checking for preemption. The next Part describes anticipated advances in technology-assisted review of law review submissions as well as the potential for harm associated with these developments.

II. THE PROMISE AND PERILS OF USING TECHNOLOGY-ASSISTED REVIEW IN THE LAW REVIEW SUBMISSIONS PROCESS

Editors may soon be able to use technology-assisted review to help screen law review submissions.[135] The usefulness, accuracy, and reliability of artificial intelligence in this context will depend on the circumstances in which it is implemented. Technology could partially automate some aspects of the existing selection process, including its potential for harm, or it could serve as a mechanism for evaluating and mitigating bias. A partially automated system might attempt to reproduce the existing submissions system if developers train it using articles published by a comprehensive set of law reviews.[136] Alternatively, developers might provide examples of articles published in law reviews that rank above a specific threshold, or articles that have been highly-cited or downloaded.[137] In a less realistic scenario, developers might enlist a group of scholars to classify the strength of articles in the training set.[138] These different approaches will be discussed in greater detail below.[139]

This Part sets forth some of the benefits of technology-assisted review of submissions, including the potential to address shortcomings in the current screening process and allow for greater anonymized review. It then details the potential harms associated with the anticipated use of artificial intelligence to help screen submissions, including the risks of

authorities with the same legal issues, facts, and jurisdiction . . . ."); LEXIS BRIEF ANALYSIS, https://plus.lexis.com/briefAnalysis (last visited May 11, 2022) [https://perma.cc/U4CF-Z4GN] ("Upload documents from both sides to see citations in common and omitted authority."); WESTLAW EDGE QUICK CHECK, https://1.next.westlaw.com/QuickCheck (last visited May 11, 2022) [https://perma.cc/2NW8-ZP9Z] ("Review the most relevant authority . . . . Discover issues with the citations and quotations relied upon by the parties.").

135 *See Home*, *supra* note 12; *Law Review System*, *supra* note 2.

136 *See infra* Part II.B.2.

137 *See infra* Part II.B.2.

138 *See infra* Part II.B.2.

139 *See infra* Part II.B.2.

enabling efficient implementation of existing bias and misplaced deference to scores assigned by algorithms.

A. *Technology Can Mitigate or Perpetuate Existing Bias in Screening*

Editors attempt to manage an overwhelming onslaught of submissions through a variety of screening mechanisms. Despite the importance of placement to an author's professional opportunities, law reviews tend not to publish their decision-making metrics and rarely disclose how they decide which submissions to accept or reject.[140] In response to surveys, however, editors have indicated factors they often consider. Some of these current practices reflect practical considerations that are unlikely to cause harm,[141] while others may result in discriminatory effects.[142] Artificial intelligence could partially automate consideration of many technical factors with minimal risk,[143] assigning a score to a given submission to indicate whether it would be likely to benefit from further review. Technology-assisted review can also facilitate anonymized review and ameliorate structural inefficiencies in the current submissions process.[144]

1. Evaluating Considerations that Are Unlikely to Cause Harm

Overworked editors need mechanisms to reduce the volume of submissions that require further examination. Some of the factors that editors assess in article selection reflect practical considerations. They are cognizant that certain pieces will require more time and effort to prepare for publication.[145] For example, they consider whether a submission complies with formatting requirements and the completeness of citations.[146] Other factors they assess include the ratio of text to footnotes, the number of footnotes, and the density of

---

[140] *See* Noah C. Chauvin, *The Banality of Law Journal Rejections*, 106 MINN. L. REV. HEADNOTES 18, 22 (2021) (observing that many journals use "template rejection notes and send essentially identical messages").

[141] *See infra* Part II.A.1.

[142] *See infra* Part II.A.2.

[143] *See* Checco et al., *supra* note 40, at 4.

[144] *See infra* Part II.A.3-4.

[145] *See* Jason P. Nance & Dylan J. Steinberg, *The Law Review Article Selection Process: Results from a National Study*, 71 ALB. L. REV. 565, 586 (2008) (stating that "Articles Editors also have an eye on the difficulty of preparing an article for publication").

[146] *See id.* at 613 (finding that a submission "that fails to conform to a journal's stylistic requirements is significantly less likely to receive an offer of publication").

footnotes.[147] Editors sometimes also examine the length of an article as a signal of quality.[148] Considering these types of factors is unlikely to result in discriminatory effects for most authors, though they may not be sound indications of quality for a given submission.[149] Still, authors without the ability to hire research assistants or other institutional support may be at a disadvantage in meeting some of these types of formalistic requirements, such as the completeness and formatting of footnotes.[150]

More substantive considerations include the timeliness of a piece, the importance of the topic, and the quality of writing and research.[151] For example, editors sometimes choose to publish articles in popular topic areas, rather than submissions that make significant contributions in narrow topics like tax law.[152] In assessing the quality of the research and writing in a submission, editors likely consider the inclusion of seminal cases, statutes, or terms of art in discussing specific areas of law.[153] They also attempt to evaluate the contribution of the paper to the literature more broadly.[154]

---

[147] *See* Arthur D. Austin, *Footnotes as Product Differentiation*, 40 VAND. L. REV. 1131, 1141-44 (1987) (describing the "density factor" and "numbers game"); Posner, *supra* note 4, at 1134; Break Into Tax, *Optimizing Law Review Submissions*, YOUTUBE (Jan. 31, 2022), https://www.youtube.com/watch?v=TZ6_eLecqQA&t=641s [https://perma.cc/664G-XNR9] (discussing the text-to-footnote ratio in an interview with former and current law review editors).

[148] *See* Lindgren, *supra* note 4, at 531 ("The extraordinary length of most legal articles is a reflection of the need to impress students."); Posner, *supra* note 4, at 1134.

[149] *See* Nance & Steinberg, *supra* note 145, at 570 n.28.

[150] *See id.*; *see also* Meera E. Deo, *Investigating Pandemic Effects on Legal Academia*, 89 FORDHAM L. REV. 2467, 2469-71 (2021) (describing the effects of the pandemic on female authors' scholarly productivity).

[151] *See* Christensen & Oseid, *supra* note 1, at 180 (observing "study participants almost unanimously agreed that they were influenced by the topic of an article"); Nance & Steinberg, *supra* note 145, at 587 (noting that "while it would appear that a small percentage of Articles Editors actively seek out trendy topics, most do not, and some assiduously avoid them").

[152] Break Into Tax, *supra* note 147; *see* Christensen & Oseid, *supra* note 1, at 196 ("Among the Top 15 segment, there was a general consensus that . . . narrow topics such as tax, civil procedure, and admiralty usually do not get published."); Nance & Steinberg, *supra* note 145, at 585 (describing how editors tend "to gravitate towards . . . articles in certain subject areas . . . as the result of a rational desire to increase the prestige of their own publications").

[153] *See* Christensen & Oseid, *supra* note 1, at 201.

[154] *See id.* at 201-03; Lindgren, *supra* note 4, at 527 (describing how student editors sometimes evaluate submissions despite a limited knowledge of the scholarly literature).

A system that partially automates examination of superficial considerations amenable to technological implementation could save editors a significant amount of time and effort.[155] Evaluating whether a submission respects formatting requirements, such as maximum word count, would be a standard application of technology that would appear to be unlikely to result in harm to most authors. Authors who lack institutional support or mentoring, however, may face greater challenges in meeting some of these requirements.[156] Law reviews, or a submissions service such as Scholastica, should offer a standardized template to authors to minimize any potential disadvantage to authors who are unaware of standard formatting.[157] At some point, a machine learning system eventually might be able to assess whether some of the formalistic aspects of a submission, such as the text-to-footnote ratio, actually correlate with features of articles that are published or highly-cited.[158] Papers may change significantly, however, between the time they are submitted and published.[159]

A partially automated system could assign a score that weights the stylistic and formatting attributes of a submission. In the context of scientific journals, similar systems have been able to predict peer review scores based on formatting, how often words were used (textual content), and the complexity of vocabulary, size of words, and the length of sentences (readability).[160] However, the ability to predict peer review results through such a topical evaluation may suggest bias exists in the peer review process.[161] As with its use in screening submissions to scientific journals, a partially automated system may not accurately assign a score to submissions in nascent areas of law or those that use unique and worthwhile writing styles.[162]

---

155 *See* Checco et al., *supra* note 40, at 4.

156 *See* Deo, *supra* note 150, at 2469-71; Nance & Steinberg, *supra* note 145, at 570 n.28.

157 *See* EUGENE VOLOKH, ACADEMIC LEGAL WRITING: LAW REVIEW ARTICLES, STUDENT NOTES, SEMINAR PAPERS, AND GETTING ON LAW REVIEW 290 (5th ed. 2016) (providing a link to a template for formatting the submission); Checco et al., *supra* note 40, at 4 (describing "first-impression bias" in document assessment).

158 *See* Berman, *supra* note 57, at 1279; *see also infra* Part III.B (describing how complete transparency about feature selection could result in strategic gaming by authors).

159 *See* Friedman, *supra* note 1, at 1317 (explaining that sometimes articles are edited and rewritten after acceptance for publication).

160 Checco et al., *supra* note 40, at 9-10.

161 *Id.*

162 *Id.*; *see also* Clements, *supra* note 31, at 4-6; Su Lin Blodgett & Brendan O'Connor, *Racial Disparity in Natural Language Processing: A Case Study of Social Media African-*

Although an algorithm's assignment of a score might appear overly objective and therefore result in too much deference by law review editors,[163] the score could provide a mechanism for evaluating bias in the current screening process.[164] Biases can exist in any decision maker, whether human or algorithmic.[165] Unlike with human editors, biases in algorithms may be encoded for later examination. It is sometimes possible to "'peer into the brain' of an algorithm"[166] and attempt to mitigate any adverse effects. Similar technology has been used in auditing employment algorithms for discrimination as well as in monitoring hiring discrimination by recruiters using online platforms.[167] For instance, if a law review editor might have passed on a submission based on dubious heuristics, a high score assigned by a machine learning algorithm might suggest implicit bias is affecting the editor's decision-making.[168] Editors could also examine how different article assignment, review, and selection approaches — for example, requiring consensus — may exacerbate or mitigate biases.

A human editor using technology-assisted review might also be able to ascertain whether a submission in a given area of law contains the most relevant citations with greater accuracy than an editor acting alone, considering the editor's limited scope of knowledge in a great number of different fields of law.[169] To discern whether relevant materials have been overlooked, a machine learning system would

American English (June 30, 2017), *in* FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY IN MACHINE LEARNING WORKSHOP AT KDD, Aug. 2017, at 1, 1 (describing how "current systems sometimes analyze the language of females and minorities more poorly than they do of whites and males").

163 *See* Heaven, *supra* note 32, at 610; *cf.* Angwin et al., *supra* note 112 (highlighting how judges have cited "risk scores" generated by algorithms "in their sentencing decisions").

164 *See generally* Cofone, *supra* note 55, at 1411 (discussing how "algorithms can be productive for reducing discrimination").

165 *See* Volokh, *supra* note 11, at 1140 (arguing the use of artificial intelligence "doesn't need to be perfect" because humans are not perfect).

166 Kroll et al., *supra* note 18, at 634; *see* Cofone, *supra* note 55, at 1411.

167 *See* Hangartner et al., *supra* note 18, at 573 (finding that contact rates by recruiters using an online platform were lower for individuals from minority groups than for members of majority groups); Alex Engler, *Auditing Employment Algorithms for Discrimination*, BROOKINGS (Mar. 12, 2021), https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination [https://perma.cc/3RMX-6W8W].

168 *See* Kleinberg et al., *supra* note 20, at 30098-100.

169 Artificial intelligence technology might be especially useful for evaluating interdisciplinary submissions. *See* Wise et al., *supra* note 8, at 11 (noting that consideration of "interdisciplinary articles requires both legal expertise and expertise in another discipline").

require a large corpus of documents to engage in related-articles analysis.[170] By comparing a submission to related articles that use similar terminology, machine learning technology can ascertain when important cases, books, or articles might have been missed.[171] For example, a partially automated system might indicate that a submission discussing abortion failed to mention *Roe v. Wade*.[172] Not only would such technology ensure the most prominent work has been cited, it also could raise the voices of marginalized authors whose work has been overlooked historically.[173]

A machine learning algorithm might be used to assign a lower score to articles that failed to cite relevant materials, though this practice could result in harm. For example, a scoring mechanism could help editors evaluate the likely completeness of research in a submission. However, authors with institutional access to the related-articles technology would have an advantage as compared with authors lacking such access.[174] Moreover, perhaps the widespread adoption of such a system might "make citations pointless" because readers could also use the related-articles technology to identify sources that are relevant to a given publication.[175]

---

170 *See* Lehr & Ohm*, supra* note 14, at 678 ("To reap the predictive benefits of machine learning, a sufficiently large number of observations is required."); Ron Snyder, *Under the Hood of Text Analyzer*, JSTOR LABS (Mar. 7, 2017), https://labs.jstor.org/blog/under-the-hood-of-text-analyzer-2 [https://perma.cc/5U5E-4DKF] (describing the technical aspects of how the Text Analyzer tool processes uploaded text, identifies the main topics and entities in it, and then suggests similar documents in the JSTOR database).

171 *See Home*, *supra* note 12 (offering a machine learning system that will "search for most on-point literature" automatically); *Text Analyzer: About*, *supra* note 16 ("The tool analyzes the text within the document to find key topics and terms used, and then uses the ones it deems most important — the 'prioritized terms' — to find similar content in JSTOR.").

172 410 U.S. 113 (1973).

173 *See Robert Anderson on Analytics for Law Review Submissions and Publishing*, *supra* note 126, at 17:00-19:07 (describing how ScholarSift has the potential to increase citation counts for people of color and women); Frye, *supra* note 127 ("Unlike other kinds of machine learning programs, which seem almost designed to reinforce unfortunate prejudices, ScholarSift seems to do the opposite, highlighting authors who might otherwise be overlooked.").

174 *See generally* Bale, *supra* note 1, at 48 (discussing the importance of institutional support in the submissions process); Kroll et al., *supra* note 18, at 657-60 (describing how systems can be strategically gamed).

175 Frye, *supra* note 127 (arguing that if the main reason for citation "is to identify relevant sources that readers will find helpful," perhaps "legal scholarship could adopt a new norm in which authors only cite works a computer wouldn't flag as relevant").

For the foreseeable future, a partially automated system will not be able to assign a score to a submission based on substantive considerations (such as novelty), other than perhaps completeness of research.[176] Artificial intelligence cannot engage in the deep, substantive analysis that is the most time-consuming aspect of submission evaluation. Nevertheless, streamlining the most straightforward elements of review would provide editors additional time to engage in more extensive examination for a greater number of submissions and serve as a feedback mechanism for editors' selection decisions.

2. Avoiding Considerations Likely to Result in Adverse Outcomes

Overwhelmed editors sometimes rely on factors that can result in adverse outcomes, even though any harmful effects are unintentional. For example, they may consider the reputation of the law school an author attended or the author's current affiliation in determining whether to review an article or make an offer of publication.[177] Some commentators have explained that the inclusion of these features can result in discrimination against minority and female applicants, even though it provides little insight into the quality of a given submission.[178] They describe how bias against non-elite schools begins during the faculty recruitment and hiring process, affecting diversity at highly-ranked schools.[179] Individuals may attend or be affiliated with less-prestigious schools for reasons that correlate with race, ethnicity,

---

[176] *See* Checco et al., *supra* note 40, at 8 tbl.5 (showcasing the potential role of AI in different aspects of the peer review process); *infra* Part III.C for a discussion of potential future applications.

[177] *See* Christensen & Oseid, *supra* note 1, at 188-91 (concluding that "law review editors, particularly those at higher ranked schools, are heavily influenced by author credentials"); Friedman, *supra* note 1, at 1315-16; Lindgren, *supra* note 4, at 530; Nance & Steinberg, *supra* note 145, at 584 (stating that a survey indicated that "editors use author credentials extensively to determine which articles to publish"); Wise et al., *supra* note 8, at 40 (noting that "two of the major criticisms of law reviews' selection practices are that law reviews frequently select articles on the basis of the author's credentials and law school affiliation rather than on article quality").

[178] *See* DEO, *supra* note 5, at 18; Barocas & Selbst, *supra* note 26, at 689; Minna J. Kotkin, *Of Authorship and Audacity: An Empirical Study of Gender Disparity and Privilege in the 'Top Ten' Law Reviews*, 31 WOMEN'S RTS. L. REP. 385, 389 (2010) ("[I]t may be that the best scholars are at the best law schools, but an effort should be made to ensure that unwarranted privilege is not at work."); Keerthana Nunna, W. Nicholson Price II & Jonathan Tietz, *Hierarchy, Race & Gender in Legal Scholarly Networks*, 75 STAN. L. REV. (forthcoming 2023) (noting that "race/gender demographics vary with school rank").

[179] *See* DEO, *supra* note 5, at 13.

religion, and gender — such as financial concerns, family obligations, or cultural preferences — notwithstanding similar abilities.[180]

Student editors may feel undue pressure from professors at their law school to agree to publish a particular submission, which may be their professor's own work.[181] Some law reviews appear to reserve a significant amount of space in their journals for in-house authors.[182] When internal authors receive priority based on their position, it prevents the publication of another submission, harming not only outside authors but also the law review itself.[183] Evidence suggests that when law reviews publish articles by internal faculty, those in-house articles receive fewer citations than articles published by outside authors.[184] Thus, law reviews may act against their own best interest in publishing articles by internal faculty, although there may be other reasons for publishing in-house articles including reputational benefits and improved relationships within the institution.[185] Favoring in-house authors also "likely has a crowding out effect" against external authors whose articles may end up being published in lower-ranked journals and may be cited less frequently than they should as a consequence.[186]

As another example, editors sometimes consider whether the author previously published in highly-ranked journals.[187] Placement history is an imperfect indication of submission quality for several reasons. Because many law reviews do not offer blind submission, the potential for "insider bias" and consideration of the author's institutional

---

180 *See id.* at 14 (describing how women and people of color do not have the financial or social support needed to move multiple times to advance their careers); GRADUATE SCH. OF EDUC., HARV. UNIV., 2017 YEAR IN REVIEW: THE COLLABORATIVE ON ACADEMIC CAREERS IN HIGHER EDUCATION 6 (2017), https://coache.gse.harvard.edu/files/gse-coache/files/coache_annual_report_2017 [https://perma.cc/CQA8-FMUL] (finding that potential opportunities for significant others may be more important considerations in employment decisions than salary).

181 *See* Adam Chilton, Justin Driver, Jonathan S. Masur & Kyle Rozema, *Assessing Affirmative Action's Diversity Rationale*, 122 COLUM. L. REV. 331, 390-91 (2022); Thomson, *supra* note 3, at 223-24 (providing statistics on the disproportionality of authors published by the journals of their home institutions).

182 *See* Thomson, *supra* note 3, at 223; Yoon, *supra* note 3, at 310.

183 *See* Yoon, *supra* note 3, at 335.

184 *See id.* at 310, 336 (concluding that "this form of protectionism creates a deadweight loss in legal scholarship"); Chilton et al., *supra* note 181, at 390-91.

185 *See* Yoon, *supra* note 3, at 310.

186 *Id.* at 335.

187 *See* Christensen & Oseid, *supra* note 1, at 180 (noting that "editors at higher tiered law schools were highly influenced by where an author has previously published").

affiliation can taint the placement process.[188] The placement process is not independent of partiality, as "disproportionate influence constructs our very notions of what good quality scholarship is."[189] Although past placement may provide some signal of the quality of a *previous* piece, past performance may not be indicative of *current* submission quality. In other words, previous article quality is no guarantee of the quality of a new submission.[190] In the scientific realm, researchers and institutions have criticized overreliance on indicators, such as journal impact factors, in assessing the quality of research.[191] They argue for improving the mechanisms for evaluation and stress the importance of considering research on its merits.[192]

Most troubling, some scholars have described how an author's gender or race may affect placement of prior work in highly-ranked law reviews.[193] Bias may influence which research is considered valuable and accepted for publication, even when it is unintentional.[194] With regard to the effects of gender, one empirical study in the area of legal studies demonstrated "significant gender disparity in publication" at the top ten law reviews.[195] Although the study analyzes many potential reasons for the discrepancy, it does not offer a conclusion as to why it

---

188 Olufunmilayo B. Arewa, Andrew P. Moriss & William D. Henderson, *Enduring Hierarchies in American Legal Education*, 89 IND. L.J. 941, 1010 (2014); *see* Lawprofblawg & Darren Bush, *Law Reviews, Citation Counts, and Twitter (Oh My!): Behind the Curtains of the Law Professor's Search for Meaning*, 50 LOY. U. CHI. L.J. 327, 364 (2018).

189 J. M. Balkin & Sanford Levinson, *How to Win Cites and Influence People*, 71 CHI.-KENT L. REV. 843, 844 (1996) (explaining that "our notions of quality are not fully separable from notions of influence, [but] not because influence necessarily follows quality as its just reward").

190 *Cf.* Lawrence Carrel, *Study Proves Past Results Don't Predict Future Results*, FORBES (Feb. 15, 2020), https://www.forbes.com/sites/lcarrel/2020/02/15/study-proves-past-results-dont-predict-future-results [https://perma.cc/7HXJ-SWNE] ("Past results are no guarantee of future performance.").

191 *See* Sugimoto, *supra* note 110, at 31; *The Declaration*, *supra* note 117.

192 *See* Sugimoto, *supra* note 110, at 31.

193 *See* Amy DeVaudreuil, *Silence at the California Law Review*, 91 CALIF. L. REV. 1183, 1187 (2003); Higdon, *supra* note 1, at 348-49; Kotkin, *supra* note 178, at 386; Leong, *supra* note 5, at 373 (finding "only 32% of law review articles are by women, and the disparity is even more significant at the 'most prestigious' law reviews, with women publishing 20.4% of articles in those venues").

194 Victor Ray, *The Racial Politics of Citation*, INSIDE HIGHER ED (Apr. 27, 2018), https://www.insidehighered.com/advice/2018/04/27/racial-exclusions-scholarly-citations-opinion [https://perma.cc/3DFT-FQSB] ("Intentionally or not, strong evidence shows that bias can inform the types of research that is considered valid and worthy of citation.").

195 Kotkin, *supra* note 178, at 386.

exists.[196] As for the effects of race on placement, one article found that almost seventy percent of articles published in the top ten law reviews in 2017 were written by authors who had graduated from the top five law schools.[197] People of color comprise a smaller proportion of graduates from top law schools, so consideration of the author's educational affiliation in screening submissions may result in discriminatory effects.[198] Status as a law professor can also affect placement in a way that causes harm to marginalized groups.[199] The available data indicates that approximately twenty-four percent of law professors are white women, eight percent are men of color, and seven percent are women of color.[200]

Until very recently, the composition of student editors at many top law reviews has tended to be fairly homogenous in terms of racial and ethnic background.[201] Some commentators have described how the lack of diversity in law review membership may affect article selection; they believe that "diverse groups of students bring different ideas about the nature of quality legal scholarship to the table."[202] Many journals have been making efforts to increase diversity among their membership.[203] Such changes are not without controversy; challenges to the diversity

[196] *Id.*

[197] Lawprofblawg & Bush, *supra* note 188, at 336.

[198] *See id.* at 336-37; Heald, *supra* note 8, at 3 (explaining that "57% of all law professors come from Harvard, Yale, Stanford, Columbia, or UChicago, and . . . 95% of all professors at the top ten schools graduated from a top ten institution").

[199] *See* Heald, *supra* note 8, at 3.

[200] Deo, *supra* note 150, at 2471.

[201] *See* STEPHANIE CHICHETTI, EMILY J. FREEBORN & LILIA VOLYNKOVA, N.Y. L. SCH. L. REV., 2011-2012 LAW REVIEW DIVERSITY REPORT 3 fig.1 (2012), https://silo.tips/download/law-review-diversity-report-5 [https://perma.cc/4GT2-HCJH] (finding that 15% of editors-in-chief of the top 50 law reviews identified as a person of color and 29% identified as female); DeVaudreuil, *supra* note 193, at 1186 (describing the consequences of having "few underrepresented students of color" on the *California Law Review* and how the journal "might begin to address the problem of the lack of diversity in its membership"); Fred R. Shapiro, *The Most-Cited Legal Scholars Revisited*, 88 U. CHI. L. REV. 1595, 1609-10 (2021) [hereinafter *Most-Cited Revisited*] (noting all the editors-in-chief at the sixteen highest-ranked law schools are female).

[202] DeVaudreuil, *supra* note 193, at 1187.

[203] *See, e.g.*, *Diversity*, CALIF. L. REV., https://www.californialawreview.org/about/diversity (last visited July 9, 2022) [https://perma.cc/6S7J-R8QV] (describing the *California Law Review's* diversity initiatives); *Writing Competition*, HARV. L. REV., https://harvardlawreview.org/writing-competition (last visited Feb. 11, 2022) [https://perma.cc/89NV-3MS7] (describing how applicants may choose to include "aspects of their identity available through the Law Review's holistic consideration process").

policies at the *Harvard Law Review* and the *New York University Law Review* have been raised, although unsuccessfully.[204]

The changing leadership at law reviews may influence the submissions that the journal selects for publication.[205] Shifts in diversity may affect group decision-making in some situations.[206] For example, one study found that companies having the greatest number of women on their boards earned a higher return on investment compared with those having the smallest number of women on their boards, though some maintain that other factors are responsible for the positive outcomes.[207] Scholars have argued that the group process of selecting law review submissions may benefit from diverse viewpoints.[208] One study found that law reviews that implemented diversity policies had "median citations to their volumes increase by roughly 23% in the ensuing five years."[209]

An artificial intelligence system should avoid replicating biases in the current submissions system. An algorithm that considers information about the law school the author attended, placement of the author's prior articles in highly-ranked journals, or the author's institutional

---

[204] Fac., Alumni, & Students Opposed to Racial Preferences v. N.Y. Univ. L. Rev., No. 18 Civ. 9184 (ER), 2020 WL 1529311, at *1 (S.D.N.Y. Mar. 31, 2020), *aff'd sub nom*. Fac. v. N.Y. Univ., 11 F.4th 68 (2d Cir. 2021); Fac., Alumni, & Students Opposed to Racial Preferences v. Harvard L. Rev., No. CV 18-12105, 2019 WL 3754023 (D. Mass. Aug. 8, 2019).

[205] *See* Karen Sloan, *'A More Diverse Conversation': Why It Matters that More Law Journals Are Electing Black Editors*, LAW.COM (Mar. 24, 2021, 2:56 PM), https://www.law.com/2021/03/24/a-more-diverse-conversation-why-it-matters-that-more-law-journals-are-electing-black-editors [https://perma.cc/6X6R-HCSB] ("The diversification of the editor-in-chief ranks may also prompt a shift in the articles and authors who get published in those journals . . . editors from diverse backgrounds are more likely to recognize the value of different perspectives and approaches within legal scholarship."); Chilton et al., *supra* note 181, at 398 (explaining that "different members of the group are able to contribute different viewpoints to the collective process" of selecting submissions for publication).

[206] *See* Chilton et al., *supra* note 181, at 360-61 (describing studies that demonstrate the impact of diversity in decision-making by juries).

[207] *Why Diversity and Inclusion Matter: Financial Performance (Appendix)*, CATALYST (June 24, 2020), https://www.catalyst.org/research/why-diversity-and-inclusion-matter-financial-performance [https://perma.cc/262Z-SR5Y]. *But see* Kim Elsesser, *What to Expect from the Influx of Women on California's Corporate Boards*, FORBES (May 21, 2021, 3:49 PM EDT), https://www.forbes.com/sites/kimelsesser/2021/05/21/what-to-expect-from-the-influx-of-women-on-californias-corporate-boards [https://perma.cc/8QQ7-LRPE] ("[T]he association between higher profits and female board members is likely due to . . . another factor like more innovative leadership strategies that positively impact both the selection of female board members and corporate profits.").

[208] *See* Chilton et al., *supra* note 181, at 398.

[209] *Id.* at 331.

affiliation may perpetuate discriminatory outcomes. The potential for harm from considering such information argues against its inclusion in a technology-assisted review system, even if it might appear to be helpful in evaluating submissions. The risks of bias in implementing a partially automated screening system, including the importance of careful feature selection and the potential for proxy discrimination, will be described in further detail below.[210]

3. Increasing the Feasibility of Anonymous Review

Anonymous review may help mitigate some of the bias that currently taints the submissions process.[211] Scholars have advocated for blind review in the submissions process,[212] but its feasibility for some law reviews has been limited by the number of submissions received each cycle.[213] Although many have argued for limiting simultaneous submissions, adopting peer review, or requiring acceptance of first offers as a way to increase the feasibility of anonymous review, these proposals have not been adopted.[214] Law review editors might be able to use technology-assisted review to help streamline consideration for some papers, freeing up time for anonymous review of a greater number of submissions.

By analogy, consider the effects of blinded auditions in the hiring of musicians in orchestras.[215] Before 1970, orchestras were essentially homogenous — most orchestra musicians were male students of a

[210] *See infra* Part II.B.3-4.

[211] *See* Higdon, *supra* note 1, at 344-49 (discussing various factors, including the author's race and gender, which can affect the selection process in non-anonymous review); Thomson, *supra* note 3, at 210 ("Legal academics generally favor a system of blind review in article selection, i.e., that the article selection process is 'blind' as to the identity of the author and in the institution(s) with which the author is affiliated.").

[212] *See, e.g.*, Friedman, *supra* note 1, at 1351-52 (proposing limits on the number of simultaneous submissions); Wise et al., *supra* note 8, at 72-73 (arguing in favor of blind peer review); *see also* Heald, *supra* note 8, at 1-3 (describing the benefits of anonymous peer review).

[213] *See* Christensen & Oseid, *supra* note 1, at 203-05 (describing challenges in managing the high volume of submissions).

[214] *See, e.g.*, Friedman, *supra* note 1, at 1352 (recognizing these proposals would be a "huge change to the culture" of law review submissions); Wise et al., *supra* note 8, at 73-74 (describing the process of article selection by adopting peer review as an "experiment").

[215] *See* Claudia Goldin & Cecilia Rouse, *Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians*, 90 AM. ECON. REV. 715, 716 (2000).

"select group of teachers."[216] When curtains were used to obscure the identity of a musician during auditions, such that the evaluators could hear but not see the musician, the likelihood that a female musician would be selected increased by twenty-five percent, according to one study.[217]

Similarly, facilitating anonymous evaluation of law review submissions would allow editors to impartially assess more papers, ensuring that worthy ideas would be recognized, regardless of an individual's status, affiliation, background, or identity.[218] To ensure objectivity throughout the submission process, any identifying information would need to remain anonymous until after a final decision had been reached.[219] Alternatively, some scholars have proposed that such information should be considered at some point in the selection process as a way to address systemic inequity or promote diversity.[220]

4. Addressing Structural Inefficiencies

The current system's reliance on students to filter through thousands of submissions each cycle is not merely inefficient,[221] it is unfair to both students and authors. Students devote numerous, unpaid hours for work that is sometimes tedious. However, they do receive many benefits

---

216 *Id.* at 715-16.

217 *Id.* at 736. *But see* Anthony Tommasini, *To Make Orchestras More Diverse, End Blind Auditions*, N.Y. TIMES (July 16, 2020), https://www.nytimes.com/2020/07/16/arts/music/blind-auditions-orchestras-race.html [https://perma.cc/2FHQ-3JHH] (arguing that orchestras should "take race and gender into account, along with the full spectrum of a musician's experience").

218 Wise et al., *supra* note 8, at 72.

219 *See* Thomson, *supra* note 3, at 226, 262 ("[I]t must be asked why the review process is blind only until the Committee's final vote and not fully blind."); *Article Submissions*, STAN. L. REV., https://www.stanfordlawreview.org/submissions/article-submissions (last visited Jan. 2, 2022) [https://perma.cc/8ERP-WJFU] ("[O]ur review process is fully blind until the Committee's final vote."); *How WLR Is Cultivating a Bias-Conscious Editorial Culture Since Implementing Blind Article Selection*, SCHOLASTICA (Aug. 28, 2020), https://blog.scholasticahq.com/post/wlr-implementing-blind-article-selection [https://perma.cc/HP8Z-XL56] (describing how the *Washington Law Review* "follows a 'partial double-blind' review process wherein author and editor identities are kept anonymous during its first two rounds of article review").

220 *See* Tommasini, *supra* note 217. *See generally* Friedman, *supra* note 1, at 1316 (describing how editors may "favor groups they worry are excluded otherwise from the publishing process, be it junior scholars, or scholars of color, or any other group they deem important").

221 *See* Bale, *supra* note 1, at 48; Christensen & Oseid, *supra* note 1, at 203-05; Friedman, *supra* note 1, at 1306-07; Higdon, *supra* note 1, at 341.

in return for their efforts: the ability to influence legal scholarship through selection and editing of articles, engagement with legal scholars, educational and research opportunities, academic credit, and potentially increased employment prospects.[222] Technology-assisted review could reduce the burden on students who screen through volumes of submissions. In the current system, some higher ranked law reviews rely on lower ranked law reviews to conduct the initial screening of submissions.[223] Elite journals might evaluate a submission in light of an expedite request based on a lower ranked journal's offer, which effectively provides a signal of quality.[224] Thus, the work of lower ranked journals is sometimes lost through the unfairness of the expedite process.[225]

Technology-assisted review would provide the greatest benefit for the journals engaged in the most intensive screening. Artificial intelligence is best used to highlight clearly deficient submissions.[226] Given the large number of law schools with general and specialty journals,[227] some journals will have a smaller number of submissions to review. Journals with more relaxed submissions standards[228] may not have as great of a need for technology-assisted review, while journals with a high volume of submissions to review will likely obtain a greater benefit from

[222] Wise et al., *supra* note 8, at 4, 24-25 ("Law review editors select and edit articles; engage in legal analysis, research, and writing; interact with legal scholars; and manage an important legal enterprise."); *see* Friedman, *supra* note 1, at 1333-34 (detailing the costs and benefits of "free student labor").

[223] Miller, *supra* note 42, at 214.

[224] *Id.*

[225] *See id.* at 214-15; Heald, *supra* note 8, at 2 (concluding the expedite process "wastes limited reviewing resources, chokes an already overwhelmed system, and creates genuine moral hazard").

[226] *See* Surden, *Machine Learning*, *supra* note 11, at 101-02 (describing how "the algorithms may be able to reliably filter out large swathes of documents that are likely to be irrelevant so that the attorney does not have to waste limited cognitive resources analyzing them").

[227] *See* Raizel Liebler, Information for Submitting to the Top Specialty Law Journals (Jan. 2020) (unpublished manuscript) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3387635) (providing a list of the top specialty journals); Rostron & Levit, *supra* note 8 (indicating that there are 196 general law reviews); *W&L Law Journal Rankings*, WASH. & LEE SCH. OF L., https://managementtools4.wlu.edu/LawJournals (last updated July 15, 2022) [https://perma.cc/R3GD-XCHV] (ranking the top 400 published law journals in the United States).

[228] *See* Friedman, *supra* note 1, at 1329 (noting that "one charm of the current system is that every article finds a home"); Heald, *supra* note 8, at 1 (explaining that with 654 law journals available, legal academics "never have to worry about getting published. It's just a question of where").

screening assistance.[229] However, if selective journals rely on technology-assisted review scores to a greater extent than less selective journals, the potential for harm increases.[230]

B. The Risks of Implementing Technology-Assisted Review

Using artificial intelligence to assist in reviewing submissions raises significant concerns. Although relying on partially automated screening to assess compliance with formatting or word count requirements is unlikely to cause harm for most authors, using artificial intelligence in a more substantive manner could result in discriminatory outcomes.[231] In their seminal work describing the risks associated with algorithmic decision-making, Solon Barocas and Andrew Selbst set forth a useful categorization to describe the ways in which bias may taint machine learning systems, which will be described in greater detail in the sections that follow.[232] The individuals involved in developing the machine learning algorithm have biases, which they may unintentionally incorporate into the technology.[233] For example, developers may specify the goal for the technology to try to match, known as the "target variable" (such as the "strength" or "weakness" of a paper), in a way that introduces unfairness.[234] They may also select training data that may not be representative or may reflect systemic bias.[235] For instance, an algorithm might consider features that embed bias — perhaps awarding lower scores to papers from authors if they did not graduate from a top three law school.[236] Algorithms may rely on factors that end up being proxies for information about race, gender, religion, or other protected characteristics of authors.[237] In these types of circumstances, technology-assisted review may perpetuate inequality in future decision-making.

229 For instance, the Duke Law Journal had an acceptance rate of 1.24% in the spring of 2021. Bale, *supra* note 1, at 48.

230 *See infra* Part II.B.

231 *See generally* Barocas & Selbst, *supra* note 26, at 677-93 (discussing multiple ways in which adverse outcomes may occur during the data mining process).

232 *Id.*

233 *Id.* at 677-81.

234 *Id.* at 678-80.

235 *Id.* at 680-81.

236 *See id.* at 689.

237 *See generally* Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1266 (2020) (defining proxy discrimination as "scenarios in which an algorithm uses a variable whose predictive power derives from its correlation with membership in the suspect class").

1. Difficulty Defining the Target Variable

Decisions made about labeling target variables can introduce bias into machine learning algorithms.[238] Target variables are the "outcomes of interest" that the technology seeks to match.[239] Discrimination can arise if developers select target variables that give some groups an advantage over others.[240] Barocas and Selbst offer the example of a hiring algorithm in which employers try to automate the process of finding a "good" employee.[241] Unlike spam detection, which is a clearly defined binary determination of "spam" or "not spam," defining what makes a "good" employee is more subjective. If developers consider prior assessments from human-evaluated annual reviews in defining a "good" employee, the definition of the target variable will be subjective and likely inconsistent.[242] Determining a "good" employee might include an examination of productivity, tenure, or sales, among other factors.[243] Discrimination can creep in during the hiring process if, for example, the algorithm considers overall tenure of employment. Because women tend to take leave from the workforce at a higher rate than men historically, a consideration of overall tenure in defining a "good" employee may result in discriminatory outcomes based on gender.[244]

In the context of law review submissions, the manner in which the developers of machine learning technology define the "quality" of an example in the training data can introduce discrimination. Sometimes, publication or the placement of a piece can provide an indication of quality.[245] As described previously, however, placement can be an unreliable signal for quality for many reasons.[246] Placement often reflects insider bias where status, including favoring of in-house faculty, influences where an article is published.[247] The author's race or gender

[238] Bent, *supra* note 55, at 811 (explaining how "the selection and labeling of the target variable creates a vulnerability in the machine-learning process that can lead to the reproduction of human bias").

[239] Barocas & Selbst, *supra* note 26, at 678.

[240] *See id.* at 679.

[241] *Id.*

[242] *Id.*

[243] *Id.*

[244] Kroll et al., *supra* note 18, at 681 (describing how the consideration of tenure is "a known proxy for gender in hiring applications").

[245] *See* Dennis J. Callahan & Neal Devins, *Law Review Article Placement: Benefit or Beauty Prize?*, 56 J. LEGAL EDUC. 374, 385 (2006) (finding that "articles in high-tier reviews continue to be cited more frequently than those published in other tiers").

[246] *See supra* Part II.A.2.

[247] *See supra* Part II.A.2.

may also adversely affect the placement of a given piece.[248] In particular, consideration of an author's educational affiliation and status as a professor may have a bearing on placement, possibly resulting in discriminatory effects.[249]

As an alternative, developers might rely on citation frequency, depth, or publication downloads to ascertain the quality of an example in the training data set.[250] A training data set that relies on these metrics could be more inclusive, as some articles from lower ranked journals receive many more citations than publications from highly-ranked law reviews.[251] Scholars have raised concerns, however, that these indicators may suffer from bias and gaming risks.[252] They also maintain that it may not be possible to remove the residual influence of

[248] *See supra* Part II.A.2.

[249] *See supra* Part II.A.2.

[250] *See* Paul J. Heald & Ted M. Sichelman, *Ranking the Academic Impact of 100 American Law Schools*, 60 JURIMETRICS J. 1, 4 (2019) (recommending consideration of citation counts and SSRN download statistics in a faculty reputation component for the U.S. News ranking score for law schools); Gregory Sisk, Nicole Catlin, Alexandra Anderson & Lauren Gunderson, *Scholarly Impact of Law School Faculties in 2021: Updating the Leiter Score Ranking for the Top Third*, 17 U. ST. THOMAS L.J. 1041, 1048 (2022) (stating that although "there are multiple ways to evaluate the scholarly work of individual law professors . . . a citation count measure is a valid and reliable proxy for scholarly excellence"); Yoon, *supra* note 3, at 314-15 (concluding that citation count "is a well-established — and the most objective — measure of quality . . . in legal scholarship").

[251] *See* Alfred L. Brophy, *The Signaling Value of Law Reviews: An Exploration of Citations and Prestige*, 36 FLA. ST. U. L. REV. 229, 236-38 (2009).

[252] *See, e.g.*, Bernard S. Black & Paul L. Caron, *Ranking Law Schools: Using SSRN to Measure Scholarly Performance*, 81 IND. L.J. 83, 122 (2006) (discussing how the number of downloads from SSRN is vulnerable to gaming); Gregory Scott Crespi, *Judicial and Law Review Citation Frequencies for Articles Published in Different "Tiers" of Law Journals: An Empirical Analysis*, 44 SANTA CLARA L. REV. 897, 901-02 (2004) (explaining "there does not appear to be any feasible way to separate out and control for relative author prestige or article quality"); Richard Delgado, *The Imperial Scholar Revisited: How to Marginalize Outsider Writing, Ten Years Later*, 140 U. PA. L. REV. 1349, 1351 (1992) (concluding at the time, that "mainstream figures who control the terms of discourse marginalize outsider writing as long as possible"); Brian Leiter, *Measuring the Academic Distinction of Law Faculties*, 29 J. LEGAL STUDS. 451, 469 (2000) ("The [cited] work is neither particularly good nor especially creative or groundbreaking, but it is there and everyone knows it is there and it must be duly acknowledged."); Deborah Jones Merritt, *Scholarly Influence in a Diverse Legal Academy: Race, Sex, and Citation Counts*, 29 J. LEGAL STUDS. 345, 347 (2000) ("[F]emale and minority scholars still lag somewhat behind white men in average citation counts. The differences, however, are small — especially when compared to other variations in citation rates . . . ."); Gregory Sisk, *Measuring Law Faculty Scholarly Impact by Citations: Reliable and Valid for Collective Faculty Ranking*, 60 JURIMETRICS J. 41, 53-54 (2019) (describing how authors have gamed SSRN downloads).

placement status on citation count.[253] In addition, articles in emerging or less well-studied areas may receive fewer citations at first.[254] With regard to the potential for discrimination, the most recent research available indicates that work authored by women receives more citations than articles written by men in the area of legal studies.[255] Additionally, the representation of women on the list of highly-cited scholars has been increasing.[256] However, further empirical investigation about the effects of race and gender on citation would be useful if citation statistics were to be used in defining the target variable of quality.[257] In addition, these types of metrics will not adequately capture the quality of a given work on an individual level.[258]

---

253 *See* Crespi, *supra* note 252, at 901-02; Jeffrey L. Harrison & Amy R. Mashburn, *Citations, Justifications, and the Troubled State of Legal Scholarship: An Empirical Study*, 3 TEX. A&M L. REV. 45, 49 (2015) (finding that the "citation of articles by law professors is highly correlated with the ranking of the review publishing the article and — in the eyes of other law professors — the prestige of the author's institutional affiliations"); *see also* Leiter, *supra* note 252, at 469.

254 *See* Fred R. Shapiro & Michelle Pearse, *The Most-Cited Law Review Articles of All Time*, 110 MICH. L. REV. 1483, 1507 (2012) (describing "subject trend[s]" in highly-cited articles); Fred R. Shapiro, *The Most-Cited Legal Scholars*, 29 J. LEGAL STUDS. 409, 413 (2000) [hereinafter *Legal Scholars*] ("Some topics have a much larger scholarly literature than others. A reasonably prolific commentator on constitutional law will have far more opportunities to be cited than even the most important writer on wills.")

255 *See* Ian Ayres & Fredrick E. Vars, *Determinants of Citations to Articles in Elite Law Reviews*, 29 J. LEGAL STUDS. 427, 427 (2000) ("[A]rticles by young, female, or minority authors are more heavily cited."); Christopher A. Cotropia & Lee Petherbridge, *Gender Disparity in Law Review Citation Rates*, 59 WM. & MARY L. REV. 771, 799 (2018) (concluding that "female-authored articles appear generally to be more cited than male-authored articles in the field of legal studies"). *But see* Delgado, *supra* note 252, at 1351; Merritt, *supra* note 252, at 347.

256 *See* Shapiro, *Most-Cited Revisited*, *supra* note 201, at 1609-10 (noting that one percent of the most-cited legal scholars are female, compared with six out of the top sixteen most-cited younger legal scholars). *But see* Deo, *supra* note 150, at 2469 (describing how the COVID-19 pandemic has adversely affected the volume of submissions by female authors); Angela Onwuachi-Willig, *The Intersectional Race and Gender Effects of the Pandemic in Legal Academia*, 72 HASTINGS L.J. 1703, 1706 (2021) (explaining that the pandemic "left women law faculty with very little of the most precious commodity needed to produce legal scholarship: time").

257 *See* Merritt, *supra* note 252, at 353.

258 *See* Shapiro & Pearse, *supra* note 254, at 1518 (discussing the limitations of citation metrics, including the lack of qualitative assessment and that a work may be cited in a negative way); Sisk, *supra* note 252, at 43 ("No single measure of faculty scholarly activity can fully capture every individual contribution. For that reason, evaluating a single professor's scholarly work requires a nuanced, multifaceted, and individually focused assessment.").

The next Subsection describes challenges in selecting the training data, including the difficulty of assessing publication quality at the individual level and constructing a representative training data set.

2. Encoding Bias in the Training Data

Choices made in selecting the training data can also lead to unfair outcomes. Machine learning algorithms use training data, which are past examples, to create models that are applied in new situations.[259] Thus, both the selection and representativeness of the training set are essential. For the former, if bias influences the composition of examples in the training set, a partially automated system will reproduce the harmful outcome.[260] With regard to representativeness, algorithms may rely too heavily on insufficient training data from which broader conclusions are drawn.[261] For instance, an artificial intelligence system for determining if a patient is having a heart attack will have a higher rate of false negatives for women if the medical records provided to the system were primarily from male patients.[262] Because machine learning makes predictions based on a limited amount of training data, deficiencies in the representative nature of the data are magnified and may harm groups that are not fairly represented in the training data.[263]

If a machine learning algorithm is trained on limited or biased data, the resulting model will not be able to provide an accurate prediction in new situations — a phenomenon known as "overfitting."[264] For instance, a hiring algorithm trained on resumes primarily submitted by male applicants might downgrade applications that include women's names on them.[265] Although screening algorithms have been marketed to companies as "decision aids" in hiring,[266] they often reject a large share of applications automatically before any human review takes

---

[259] Surden, *Machine Learning*, *supra* note 11, at 91-94; *see* Barocas & Selbst, *supra* note 26, at 680-81 (defining "training data" as "the data that train the model to behave in a certain way").

[260] Barocas & Selbst, *supra* note 26, at 681-83 (referring to this as a "garbage-in-garbage-out" problem).

[261] *Id.* at 688-90; IBM Cloud Education, *Overfitting*, IBM (Mar. 3, 2021), https://www.ibm.com/cloud/learn/overfitting [https://perma.cc/9HBA-5HDM].

[262] *See* Suresh & Guttag, *supra* note 38, at 1.

[263] Barocas & Selbst, *supra* note 26, at 686.

[264] IBM Cloud Education, *supra* note 261 (defining overfitting as the situation "when a statistical model fits exactly against its training data" such that "the algorithm unfortunately cannot perform accurately against unseen data").

[265] *See* Dastin, *supra* note 27.

[266] Bogen, *supra* note 24.

place.[267] Problematically, hiring algorithms have based their decisions on past practices, which often contain significant biases.[268] For example, Amazon developed a hiring tool using artificial intelligence in an attempt to screen resumes in an efficient manner.[269] The programming team taught the algorithms to identify 50,000 terms that had been used in applicants' resumes in the past.[270] The technology ended up allocating minimal weight to frequently-used terms, such as various programming languages that were common among applicants.[271] However, the tool preferred words that male applicants tended to use more often, such as "executed."[272] It also learned to penalize resumes that indicated attendance at certain all-female universities.[273] Similarly, search technology and translation tools have provided results that reflect gender bias in training data, such as associating "nurses" with being female and "CEOs" with being male.[274] A lack of representative training data has undercut the usefulness and inclusiveness of many other types of machine learning technologies, resulting in reputational and social harms.[275]

---

267 *Id.*

268 *Id.*

269 *See* Dastin, *supra* note 27.

270 *Id.* (noting that the training data was sourced from applications that Amazon received, with a majority coming from male applicants).

271 *Id.*

272 *Id.*

273 *Id.*

274 *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 93-94 (2019) (describing how "only 11 percent of the top 100 'CEO' image search results from Google included women, even though 27 percent of CEOs in the United States are women"); Matthew Kay, Cynthia Matuszek & Sean A. Munson, *Unequal Representation and Gender Stereotypes in Image Search Results for Occupations*, CHI '15: PROC. 33D ANN. ACM CONF. ON HUM. FACTORS COMPUTING SYS. 3819 (2015), https://doi.org/10.1145/2702123.2702520 [https://perma.cc/BA55-R6AR] (describing gender-biased results in image searches for occupations); Calo, *supra* note 10, at 411-12 (discussing "problems involving the design and deployment" of artificial intelligence).

275 *See, e.g.*, Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 1, 1-2 (2018), https://proceedings.mlr.press/v81/buolamwini18a.html [https://perma.cc/A2UE-YTE2] (describing how image recognition technology that was trained using data consisting of primarily Caucasian faces has struggled to recognize diverse individuals); PATRICK GROTHER, MEI NGAN & KAYEE HANAOKA, NAT'L INST. OF STANDARDS & TECH. INTERAGENCY OR INTERNAL REP. NO. 8280, FACE RECOGNITION VENDOR TEST (FRVT) PART 3: DEMOGRAPHIC EFFECTS 2 (2019), https://doi.org/10.6028/NIST.IR.8280 [https://perma.cc/62MS-2PCK] (concluding that facial recognition technology was

In training an algorithm in the context of law review submissions, different options exist for selecting examples to constitute the training set. Developers might attempt to replicate the current law review submissions system by using articles previously published by a broad selection of journals as the training data.[276] In searching for related articles, JSTOR and ScholarSift both appear to rely upon vast databases of published law review articles, with JSTOR including interdisciplinary and primary materials in a variety of other fields as well.[277]

Alternatively, developers might use publications from journals ranked above a specified threshold. For instance, developers might decide to select examples from articles published by the top fifty law reviews. To the extent the placement of previous submissions incorporated bias, however, technology-assisted review would simply enable efficient implementation of past harm.[278] Deficient selection of the examples used to train the algorithm may negatively impact groups that are not adequately represented in the training data.[279] As described in the previous Subsection, using placement of a publication in a highly-ranked journal as an indication of quality could reflect insider influence and perpetuate discrimination based on race or gender.[280] In another approach, the training set could also include highly-cited or downloaded articles, regardless of the rank of the law review where they were published.[281] However, citation and download metrics may suffer from bias and gaming risks, and they may not be accurate in assessing quality on an individual level.[282]

In a less feasible scenario, developers might wish to train the algorithm based on past examples of law review articles that a group of

between 10 and 100 times more likely to misidentify Asian and Black people than white people).

[276] *See* Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 619-30 (2018) (describing how the fair use doctrine could protect the use of copyrighted works in training machine learning systems); Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. 1607, 1608 (2009) (examining whether "a nonexpressive use, which nonetheless requires copying the entirety of a copyrighted work, [can] be found to infringe the exclusive rights of the copyright owner").

[277] Shucha, *supra* note 127, at 103, 107; *Home*, *supra* note 12; *Text Analyzer: About*, *supra* note 16; *What's in JSTOR*, JSTOR, https://about.jstor.org/whats-in-jstor (last visited Apr. 21, 2022) [https://perma.cc/Y462-4QET] (describing how it combines "scholarship and primary sources on one platform").

[278] *See* Cofone, *supra* note 55, at 1398.

[279] *See* Barocas & Selbst, *supra* note 26, at 684-86.

[280] *See supra* Part II.A.2.

[281] *See supra* Part II.B.1.

[282] *See supra* Part II.B.1.

scholars classified as having a specified level of quality. This method of constructing the training data set seems unworkable for many reasons. Most importantly, it seems nearly impossible to obtain sufficient and representative involvement by legal scholars to develop training data with an acceptable number of examples of strong articles in a variety of legal fields to create a reliable predictive model.[283] In addition, legal scholars may not be able to achieve consensus on the definition of a quality publication.[284] The team also might not be able to agree on whether to produce a training data set that reflects the pool of quality publications as it currently exists or what it might ideally include to counterbalance social biases.[285] By analogy, in the context of hiring algorithms, some scholars have argued for constructing a data set to reflect how the pool might look if it did not have systemic bias.[286] For all these reasons, this last alternative way of developing the training data set is the least realistic.

The training data would need to contain sufficient examples from different areas of law, which may be problematic if inclusion in the training data depends primarily on articles with high citation counts or downloads. In areas of law that are less well developed, the data that is available to train algorithms may be incomplete or not representative.[287] If the training data comprises insufficient articles in certain fields of study, like admiralty, the machine learning technology will not be able to accurately evaluate the thoroughness of research, for example, in

283 *See* Heaven, *supra* note 32, at 610.

284 *See* Surden, *Artificial Intelligence and the Law*, *supra* note 23, at 1308-09, 1322-25 (describing how artificial intelligence "tends to work best for activities where there are underlying patterns, rules, definitive right answers, and semi-formal or formal structures that make up the process" as opposed to "areas that are conceptual, abstract, value-laden, open-ended, policy- or judgment-oriented").

285 Bent, *supra* note 55, at 807; Cofone, *supra* note 55, at 1424; *see also* Bornstein, *supra* note 22, at 541-44, 550 (describing "antisubordination theory" and introducing an "antistereotyping approach" to making algorithms antidiscriminatory).

286 *See* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1041 (2017) (explaining that "the decisionmaker must take race and gender into account in order to ensure the fairness of the result"); Cofone, *supra* note 55, at 1424; Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 887 (2017) (describing algorithmic bias that "coincides with systematic disadvantage to protected classes").

287 *See* Balkin & Levinson, *supra* note 189, at 843-44; Barocas & Selbst, *supra* note 26, at 688-90; Higdon, *supra* note 1, at 348-49; *see also* Volokh, *supra* note 11, at 1168 ("If this training data contains biases (for example, imagine a criminal trial data set in which the black defendants were convicted 95% of the time but the white defendants only 75% of the time), the AI's learning process may incorporate those biases.").

some submissions. When machine learning technology is trained on articles from a narrow group of law reviews or from a limited perspective, the system will not consistently and accurately assign scores to certain types of submissions.[288]

The model used in scoring submissions would need to be tested for validity and reliability. During training, some portion of the identified set of examples would need to be cordoned off to use later for testing the system.[289] The test data would be used to assess if the machine learning system was appropriately scoring articles.[290] If the machine learning system did not appropriately score articles in the test data, the developers would seek to understand why there was a discrepancy.[291]

3. Feature Selection and Systemic Bias

Even if a machine learning algorithm is trained with representative data, it may still reflect embedded systemic bias.[292] To mitigate such bias, developers can decide which features to include for consideration by a machine learning system. Features are the "observed variables" that the algorithm is permitted to access in detecting patterns.[293] Decisions as to which features to include can result in unfair outcomes.[294] Because a model cannot adequately represent the complexity of every individual situation, some groups will be affected by "statistically sound inferences that are nevertheless inaccurate."[295]

As an example, hiring algorithms tend to place undue emphasis on the reputation of the educational institutions that applicants attended.[296] Decisions that rely on this information will systemically overlook applicants from protected groups if equivalently competent applicants graduate from higher-ranked universities at lower rates.[297] Barocas and Selbst have analogized the overemphasis on educational reputation in hiring algorithms to "redlining."[298] Redlining is an unethical and illegal practice by which financial institutions used broad

288 *See* Balkin & Levinson, *supra* note 189, at 843-44; Barocas & Selbst, *supra* note 26, at 688-90; Higdon, *supra* note 1, at 348-49.

289 *See* Lehr & Ohm, *supra* note 14, at 698 (describing the process of model training).

290 *See id.*

291 *See id.*

292 Cofone, *supra* note 55, at 1404-05.

293 Bent, *supra* note 55, at 813.

294 *See* Barocas & Selbst, *supra* note 26, at 684-86.

295 *Id.* at 688.

296 *See id.* at 689.

297 *See id.*

298 *Id.*

criteria, such as neighborhoods, to make distinctions between subgroups in determining whether racial minorities could obtain a mortgage.[299] Barocas and Selbst argue that overemphasizing the reputation of the institution from which an applicant graduated provides minimal insight into an individual's capability.[300] Although other information could allow for a more accurate and fairer determination, employers (or hiring algorithms) may choose to focus on educational reputation because it is cost-efficient to rely on such readily available information.[301]

A partially automated screening algorithm for submissions that places too much weight on the reputation of the author's institutional affiliation or the law school they attended could result in discriminatory effects against minority and female authors, even though any adverse outcomes are not intentional.[302] Scholars have documented issues related to the lack of diversity at elite schools.[303] Despite having similar capabilities, individuals may be affiliated with or graduate from less-prestigious institutions for reasons correlated with gender, race, religion, and ethnicity.[304] Commentators have argued that such information tends not to provide a useful indication of competence at the individual level.[305] Biases related to the consideration of educational reputation or institutional affiliation that currently exist in law review article selection could be replicated by including such features in a machine learning algorithm.[306]

4. Problematic Proxies

Machine learning systems can also discriminate through the use of proxies. Although developers might instruct an algorithm to ignore gender, race, or other protected characteristics, the removed characteristic can often be ascertained through related proxies, a

[299] *See* RICHARD ROTHSTEIN, THE COLOR OF LAW: A FORGOTTEN HISTORY OF HOW OUR GOVERNMENT SEGREGATED AMERICA, at vii, 64-67 (2017) ("Banks discriminated with 'redlining,' refusing to give mortgages to African Americans . . . ."); Barocas & Selbst, *supra* note 26, at 689.

[300] Barocas & Selbst, *supra* note 26, at 689.

[301] *Id.*

[302] *See supra* Part II.A.2.

[303] *See* DEO, *supra* note 5, at 13.

[304] *See id.* at 14; GRADUATE SCH. OF EDUC., *supra* note 180.

[305] *See* DEO, *supra* note 5, at 18; Barocas & Selbst, *supra* note 26, at 689.

[306] *See supra* Part II.A.2.

phenomenon known as "redundant encoding."[307] For example, although gender might be excluded from a hiring algorithm, the overall tenure of employment could serve as a proxy for gender, if a greater number of women than men on average take leave from the workforce to have children.[308] Consequently, removing the protected characteristic as a variable for the algorithm to consider does not preclude the algorithm from considering related correlations, which can result in discriminatory effects.[309]

An algorithm that excluded protected characteristics from consideration in screening law review submissions might still raise concerns. Citations to articles from certain areas of law, such as critical race theory, or even to diverse authors might be a proxy for race.[310] An algorithm that assigned a lower score to the use of certain phrases might reproduce existing biases.[311] Considering the law school that the author attended in a screening algorithm could redundantly encode protected characteristics in a way that allows for algorithmic discrimination on the basis of race, gender, or religion.[312] Similarly, acknowledgments in an author's biographical footnote can serve as an indication of quality, but they may also reveal "potential entrenchment of existing academic privilege" or serve as a proxy for race or gender.[313]

---

307 Prince & Schwarcz, *supra* note 237, at 1275 (concluding that "AIs can and will use training data to derive less intuitive proxies for directly predictive characteristics when they are deprived of direct data on these characteristics due to legal prohibitions").

308 Kroll et al., *supra* note 18, at 681 (explaining that "women who leave a job to have children lower the average job tenure for all women, causing this metric to be a known proxy for gender in hiring applications").

309 *See* Bent, *supra* note 55, at 816 (describing "approaches to algorithmic fairness . . . [that] take protected characteristics into account"); Prince & Schwarcz, *supra* note 237, at 1266-67 (explaining how "laws that prohibit discrimination based on directly predictive characteristics must adapt to combat proxy discrimination" and providing "a menu of potential strategies" to ascertain whether artificial intelligence is engaging in proxy discrimination).

310 *See* Nunna et al., *supra* note 178; Tietz & Price, *supra* note 4, at 346.

311 *See* Barocas & Selbst, *supra* note 26, at 679-80 (describing how "different choices for the target variable . . . may have a greater or lesser adverse impact on protected classes"); Bent, *supra* note 55, at 811 (explaining that "the way the user defines and assigns a specific value to the target variable, if correlated with a protected characteristic, could unintentionally trigger a disparate impact"); Clements, *supra* note 31, at 4-6.

312 *See* Prince & Schwarcz, *supra* note 237, at 1272 (explaining that "proxy discrimination can be either intentional or unintentional").

313 Tietz & Price, *supra* note 4, at 346; *see also* Nunna et al., *supra* note 178 (finding that "authors tend to acknowledge scholars from peer schools, most of all their own school, but also to typically acknowledge folks from somewhat fancier schools" and

Although there are significant risks associated with implementing technology-assisted review of law review submissions, the next Section will describe how to examine its potential benefits and harms as compared with the current submissions process.

C. Not Letting Perfection Be the Enemy of the Good

Artificial intelligence does not have to be perfect to be useful, though developers should continue to improve upon it. If an autonomous vehicle can perform at least as well as a human driver, it is a worthwhile technology, even if it needs to be refined.[314] The use of technology-assisted review may provide superior outcomes to human editors acting alone. By analogy, Stanford University recently offered automated feedback to students using artificial intelligence in one of its online computer programming courses.[315] In online education, a course may have thousands of students, so instructors may not be able to provide the ideal amount of feedback to students.[316] During the online course at Stanford, the automated system gave 16,000 instances of feedback, with which students agreed 97.9 percent of the time.[317] Surprisingly, students agreed with the feedback they received from human instructors less often — only 96.7 percent of the time.[318] The use of automated scoring is not well suited, however, for "original research pieces."[319]

As another example, Google Translate offers automated translation where resources for translation might not otherwise exist.[320] It would be unwise to depend on Google Translate for important decisions, like understanding a plea agreement in a different language. However, Google Translate can be useful where one might not otherwise have access to translation tools, such as when communicating with a local resident.[321] Similarly, if technology-assisted review allows for additional

"men are acknowledged more than women and nonbinary scholars, and white scholars more than scholars of color").

[314] *See* Volokh, *supra* note 11, at 1139 (noting that "ordinary drivers don't set that high a bar").

[315] Metz, *supra* note 40.

[316] *Id.*

[317] *Id.*

[318] *Id.*

[319] Stephen P. Balfour, *Assessing Writing in MOOCs: Automated Essay Scoring and Calibrated Peer Review*, 8 RSCH. & PRAC. ASSESSMENT 40, 46 (2013).

[320] Surden, *Machine Learning*, *supra* note 11, at 100; *Translation AI, supra* note 40.

[321] Surden, *Machine Learning*, *supra* note 11, at 100; *Translation AI*, *supra* note 40.

opportunities for authors facing other heuristic biases,[322] it may still be worth using in limited circumstances and with appropriate oversight.

Results from a recent empirical study suggest that some authors might even prefer technology-assisted review to the current selection process.[323] The study demonstrated that users have a significant preference for automated decision-making when it provides "benefits in speed, cost, or accuracy."[324] Although they have a "mild" preference for human decision-making when the stakes increase, their preference can be overcome by "more concrete considerations, such as speed or cost, and by the default setting."[325] A technology-assisted review system for streamlining the law review submissions process would not automate decision-making; it still requires a human editor to evaluate a submission. Nevertheless, the study's results seem to support the idea that some authors might prefer a partially automated system based on the benefits it would offer in terms of speed and perhaps accuracy[326] — at least for authors confronting biases.

Despite these possible benefits, partial automation of the law review submissions process may simply enable efficient implementation of flawed practices without the possibility of "technological due process" for authors.[327] The next Part will describe how some of the potential harms associated with technology-assisted review of submissions might be addressed.

III. OVERSEEING ALGORITHMS IN IMPLEMENTATION

Implementing an artificial intelligence system to streamline the law review submissions process will be time-consuming, costly, and involve risks.[328] The training data will need to include vast amounts of publications from a multitude of different subject areas. Although law

---

[322] *See* Volokh, *supra* note 11, at 1140 (describing how the use of artificial intelligence can help with both efficiency and accessibility).

[323] *See* Derek E. Bambauer & Michael Risch, *Worse Than Human?*, 53 ARIZ. ST. L.J. 1091, 1094-97 (2022).

[324] *Id.* at 1094.

[325] *Id.*

[326] *See id.*

[327] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 5, 8 (2014) ("There is nothing unbiased about scoring systems.").

[328] *See* Andrew Ng, *AI Doesn't Have to Be Too Complicated or Expensive for Your Business*, HARV. BUS. REV. (July 29, 2021), https://hbr.org/2021/07/ai-doesnt-have-to-be-too-complicated-or-expensive-for-your-business [https://perma.cc/6BRH-D5D2] (describing how "the economics of an individual project might not support hiring a large, dedicated AI team").

schools and journals will benefit from the system, they may lack the financial resources or expertise to develop individualized artificial intelligence systems.[329] Instead, private companies will likely develop the partially automated screening technology coupled with their own submissions system or with the goal of selling the technology to other submissions services like Scholastica.[330]

Artificial intelligence systems can exacerbate underlying unfairness, as they enable bias to be embedded consistently in a system.[331] Mechanisms should be instituted to address some of the potential harms previously discussed. To the extent that racism, sexism, or other types of bias that taint the current selection process are replicated in a partially automated system, attempts to debias the system would merely be performative.[332] By incorporating "impartiality by design" and "impartiality by testing" mechanisms, developers could implement measures to minimize the risks of harm.[333] The earlier Sections of this Article describe ways to design the system to reduce the potential for adverse outcomes, including ensuring careful feature selection and representative training data.[334] For instance, developers could set up the system to disregard specific attributes, such as the author's name, to prevent gender or racial bias. However, some commentators have argued for the consideration of such information to ensure fairness.[335]

---

329 *See Robert Anderson on Analytics for Law Review Submissions and Publishing*, *supra* note 126, at 24:34-28:05.

330 *See id.* at 3:33-5:40, 24:34-28:05.

331 *See* Cofone, *supra* note 55, at 1398 ("Automated decision-making . . . brings perfect consistency across decisions."); Pauline T. Kim, *Big Data and Artificial Intelligence: New Challenges for Workplace Equality*, 57 U. LOUISVILLE L. REV. 313, 321-22 (2019) [hereinafter *Big Data and AI*] (describing how an algorithm that "makes predictions across cases or populations in a way that is *systematically* wrong or biased . . . raises much broader social concerns").

332 *See* Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 191 (2017) (arguing that "the causes of bias often lie not in the code, but in broader social processes"); Kim, *Big Data and AI*, *supra* note 331, at 320 ("If the employer's prior hiring practices excluded certain groups . . . the algorithm will simply reproduce the previously existing biases."); Kroll et al., *supra* note 18, at 681 (concluding that "if discrimination is already systemic, new data will retain the discriminatory impact"); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2251 (2019) (explaining "what prediction does is identify patterns in past data and offer them as projections about future events").

333 Volokh, *supra* note 11, at 1168-69.

334 *See supra* Part II.B.

335 *See* Bent, *supra* note 55, at 807 (stating that "the best way to get fair algorithmic results is not by hiding the protected trait, but instead by *using* the protected trait to set a fairness constraint within the algorithm design"); Chander, *supra* note 286, at 1041

For impartiality by testing, developers or an oversight organization could test for "potentially prejudiced emergent properties" and work to address unfairness as they observe it.[336] Returning to the example of Amazon's hiring algorithm, developers realized that the tool ended up teaching itself to downgrade resumes that mentioned women's colleges.[337] Although Amazon decided to discard its hiring tool as a result,[338] developers might be able to adjust the system in some circumstances to mitigate any harmful effects.[339]

The discussion below will summarize possible oversight measures to address bias in implementing an artificial intelligence system to help review law journal submissions. It will discuss the importance of regularly auditing the outcomes of a technology-assisted review system, as well as the benefits and limits of transparency. In some circumstances, developers can design a system to ensure greater accountability than transparency alone would accomplish.[340] This Part will also set forth potential future applications of artificial intelligence in the law review submissions process.

A. Regular Auditing

Auditing is the independent assessment of whether a system conforms to applicable standards and procedures, as well as to discover any interference with the operation of the system.[341] A consortium of law reviews, or a group like the Association of American Law Schools ("AALS"), could require independent oversight and regular auditing of algorithmic systems used in technology-assisted review of submissions. The oversight group should consist of a team diverse in thought, demographics, and background to oversee regular audits of the technology. Similar to obligations imposed on employers in auditing

(arguing for the inclusion of race and gender in algorithmic decision-making to reduce unfairness); Kroll et al., *supra* note 18, at 685 ("Blindness to a sensitive attribute has long been recognized as an insufficient approach to making a process fair."); Prince & Schwarcz, *supra* note 237, at 1302-03 (explaining that prohibiting consideration of protected characteristics "may effectively prevent traditional intentional proxy discrimination," but artificial intelligence "will inevitably identify other proxy variables for directly predictive data").

336 Volokh, *supra* note 11, at 1169.

337 *See* Dastin, *supra* note 27.

338 *Id.*

339 *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1310 (2008) ("Rigorous testing reflects a norm of proper software development.").

340 Kroll et al., *supra* note 18, at 637.

341 *Id.* at 660-61.

their hiring algorithms,[342] the oversight group would require inspection of partially automated systems to protect against adverse impacts against certain groups and debias screening algorithms as much as possible.[343] It would confirm that the technology-assisted review system was reliably and consistently assigning scores.[344] The group would also engage in quality assessment on an ongoing basis to minimize the likelihood that the technology perpetuates bias.[345]

Developers should design the system in a way that allows for accountability, enabling oversight to ensure that any specified rules have been applied consistently.[346] In addition to ensuring "procedural regularity"[347] in the system's implementation, an oversight group should also evaluate whether the rules used in screening are justified. The group would require analysis of data related to the submissions to determine if bias might exist in the algorithm.[348] One advantage of some

[342] *See, e.g.*, UNIF. GUIDELINES ON EMP. SELECTION PROCS. (EQUAL EMPL. OPPORTUNITY COMM'N 2021), https://www.uniformguidelines.com/uniformguidelines.html#20 (last visited July 25, 2021) [https://perma.cc/E5TL-8K67] (setting forth "a framework for determining the proper use of . . . selection procedures" in employment, including standards for validity studies); *see also* Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 659-73 (2021) (describing the benefits of mandating internal and external audits in the use of hiring algorithms).

[343] *See* Barocas & Selbst, *supra* note 26, at 715-16 (suggesting that various data points be tested to reduce disparate impact while still maintaining accuracy); Citron & Pasquale, *supra* note 327, at 18-30 (stressing the importance of auditing for scoring systems to mitigate potential harm); Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. (forthcoming 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4066781 [https://perma.cc/RG99-AH57] ("Accuracy is a critical factor in evaluating the utility of any decisionmaking system. However, an emphasis on accuracy brings its own complexity. False and true positives and negatives often differ in seriousness . . . ."); Kroll et al., *supra* note 18, at 695-705 (describing how technological tools can be used to mitigate unfairness in automated decision-making and suggesting the need for computer scientists and law makers to work together in addressing bias).

[344] *See* Lehr & Ohm, *supra* note 14, at 698.

[345] *See, e.g.*, Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 118-19, 168-82 (2017) (proposing that companies provide "algorithmic impact statements" to disclose the anticipated effectiveness and possible disparate impact of a given technology and possible alternatives); DILLON REISMAN, JASON SCHULTZ, KATE CRAWFORD & MEREDITH WHITTAKER, ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY 4 (2018), https://ainowinstitute.org/aiareport2018.pdf [https://perma.cc/LU23-ZSAQ] (describing the key elements of an algorithmic impact assessment).

[346] Kroll et al., *supra* note 18, at 637.

[347] *Id.* at 656-57.

[348] *See* Citron, *supra* note 339, at 1310-11.

types of artificial intelligence systems over human decision makers is that biases and "faulty logic" are "literally coded" in algorithms.[349] Subsequent reviewers can sometimes detect bias, whether conscious or unconscious, when auditing the system.[350] Enabling this type of oversight would require journals to collect demographic information from authors, preferably on a voluntary basis with consent obtained for the use of the information in this manner. Journals should advise authors if their information will be used for any other purpose, including for screening. For example, research might discover that the algorithm has relied on features correlated with race that result in lower scores assigned to papers written by people of color. Such revelations should cause developers to examine the technology to understand why the imbalance occurred.[351]

Technical approaches might alleviate some of the risks of stereotypes inherent in the natural language processing used by machine learning models.[352] For example, researchers have proposed a technique for developing gender-neutral models without sacrificing functionality.[353] Their proposed solution identified gender-neutral words while using word vectors that represent meanings of the word.[354] Using the researchers' model, the word "programmer" would be gender-neutral by definition.[355] In contrast, for a standard training model, the word "programmer" is associated more closely with "male" than "female."[356] Ideally, similar technical approaches would be incorporated in the design of the partially automated system, and they might also be used in response to disparities revealed through auditing. Author-related attributes could be included as part of the data to train the machine learning models to avoid bias and allow for auditing.[357] However,

[349] Cofone, *supra* note 55, at 1411.

[350] *See id.*

[351] *See id.*

[352] *See* Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 5 (2017) (describing technical approaches to address the risks of stereotypes inherent in natural language processing).

[353] JIEYU ZHAO, YICHAO ZHOU, ZEYU LI, WEI WANG & KAI-WEI CHANG, LEARNING GENDER-NEUTRAL WORD EMBEDDINGS 4847 (2018), https://arxiv.org/pdf/1809.01496.pdf [https://perma.cc/8F3R-ALX7].

[354] *Id.*

[355] *Id.*

[356] *Id.*

[357] *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1130-37 (2018).

technical approaches to mitigating algorithmic bias can be challenging, incomplete, and costly.[358]

B. Transparency and Its Limits

Some measure of oversight can be ensured through transparency.[359] The developers of the machine learning system should indicate how they selected the oversight team and its composition. They should describe how they selected the training data set. For example, developers could indicate if they are relying on publications from the top fifty journals, highly-cited articles, or some combination. Developers should also indicate which author attributes the algorithm considers or will not consider, such as gender, race, or institutional affiliation.

Law reviews should be more transparent as well. They should disclose how many submissions they receive and how many offers they make each cycle, as well as the degree to which they have implemented a partially automated system. Regular reports of publication offer rates among law reviews could be useful information for authors, regardless of whether the law review editors are using a partially automated screening system. If a law review does use artificial intelligence technology, it should also release information about how it uses the data about each submission. For example, law reviews should describe if they provide a review of all submissions regardless of their assigned scores or only those submissions that receive a score above a specified threshold. They should specify to what extent they engage in anonymous review of submissions.

Editors could agree to a code of ethics that would set forth measures taken to ensure the use of machine learning technology does not result in unfair screening.[360] They should examine whether the stated goals of their journals align with the implementation of partially automated screening software. Editors might commit to providing double-blind

---

[358] *See* Barocas & Selbst, *supra* note 26, at 716-19.

[359] *See* Citron & Pasquale, *supra* note 327, at 24-25 (explaining that scoring systems should be transparent in light of their potential for harm); Selbst & Barocas, *supra* note 357, at 1087-88 (maintaining that transparency in automated decision-making should include a description of how a model has been developed). *But see* Kroll et al., *supra* note 18, at 657-60 (concluding that "it is often necessary to keep secret the elements of a decision policy, the computer systems that implement it, key inputs, or the outcome" to "prevent strategic 'gaming' of a system").

[360] *See generally* Katyal, *supra* note 274, at 108-15 (recommending that computer scientists and software engineers adopt codes of conduct).

review for a feasible number of articles each submissions cycle, perhaps those that have received above a specified score by the screening algorithm.[361] However, editors should be cognizant of the limitations of algorithms in the submissions process and not defer too greatly to scores assigned by the system. Similar to the classification of documents that are likely irrelevant in litigation document review described previously,[362] partially automated screening in the context of law review submissions will be much better at identifying clearly deficient submissions than discerning stronger ones.[363] To mitigate potential harm, editors could also agree to a random review of a certain number of articles that were assigned low scores by the technology.[364] The random selection would require that the reviewing editors were not aware that the set of articles had been assigned low scores by the screening tool. This ongoing randomized review and assessment would help ensure reliability and identify potential bias.[365] To assist authors without institutional support, law reviews or submission services should also provide a template for authors who may be unaware of standardized formatting as well as a no-cost submission method.[366]

Providing complete transparency about the partially automated screening system would be problematic. The data or technology used in the system may be protected by intellectual property law.[367] In addition,

361 *See* Rostron & Levit, *supra* note 8; Thomson, *supra* note 3, at 226, 262.

362 *See supra* Part I.A; Surden, *Machine Learning*, *supra* note 11, at 101-02.

363 *See* Surden, *Machine Learning*, *supra* note 11, at 98.

364 *See* Kroll et al., *supra* note 18, at 639 (explaining that "while transparency of a rule makes reviewing the basis of decisions more possible, it is not a substitute for individualized review of particular decisions").

365 *See id.* at 684 (stating that "if the algorithm is designed to incorporate an element of randomness . . . the validity of the initial assumptions can be tested and the accuracy and fairness of the entire system will benefit over time"); *see also* Citron & Pasquale, *supra* note 327, at 18 (describing how "scoring systems have the potential to take a life of their own, contributing to or creating the situation they claim to merely predict").

366 *See* VOLOKH, *supra* note 157, at 290 (providing a link to a template for formatting the submission).

367 *See* Clark D. Asay, *Artificial Stupidity*, 61 WM. & MARY L. REV. 1187, 1194-97 (2020) (examining how intellectual property rights affect the development of artificial intelligence systems); Jeanne C. Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 N.Y.U. L. REV. 706, 708 (2019) (arguing that the use of trade secret to protect machine learning technology is excessive); Kroll et al., *supra* note 18, at 639 ("[D]isclosure of the data may be undesirable or even legally barred."); Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 757-58 (2021) (discussing "copyright in the individual components of the database" used to train machine learning systems); Levendowski, *supra* note 276, at 619-30 (discussing how invoking fair use might create fairer AI systems); Brenda M. Simon & Ted Sichelman, *Data-Generating Patents*, 111 NW. U. L. REV. 377, 383 (2017)

some aspects of the technology or data may affect the proprietary interests of the creator or users of the algorithm.[368] In some circumstances, regulations or statutes prevent disclosure of certain types of data.[369] Additionally, machine learning technology is not always able to provide a complete explanation of its decision-making process.[370] As new articles are published and the machine learning technology updates its system, developers might not be able to explain the reasoning behind changes to the scoring mechanism.[371]

Even when explanation is possible, complete disclosure of the details of the partially automated screening system might lead some authors to tailor their submissions.[372] Authors would be more likely to tweak variables that are easy to adjust, such as the length of an article or the footnote-to-text ratio, than those that are difficult to alter, such as the analysis of pertinent scholarship.[373] Even without complete disclosure, some savvy individuals might be able to reverse engineer the system to

(describing the use of patents and trade secrets as complementary forms of protection for data-generating inventions).

368 *See* Kroll et al., *supra* note 18, at 639 ("[D]isclosure of the data may be undesirable or even legally barred."); W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775, 788 (2021) ("[A] machine-learning developer may, for reasons of competitive advantage, want to maintain secrecy over one or more of the following aspects of its work product: the learning algorithm's source code, associated parameters, the training data, training process, or the resulting model.").

369 *See, e.g.*, Family Educational Rights and Privacy Act of 1974 (FERPA), 20 U.S.C. § 1232g (2018) (law regulating the privacy of educational records).

370 *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 6-17 (2015) (describing the secrecy and complexity of algorithms); Citron & Pasquale, *supra* note 327, at 6 (defining black boxes); Desai & Kroll, *supra* note 352, at 5 (explaining that "fundamental limitations on the analysis of software meaningfully limit the interpretability of even full disclosures of software source code"); Kroll et al., *supra* note 18, at 658 (explaining that sometimes "the purpose of the automated decision process is to determine something not directly measurable"); Selbst & Barocas, *supra* note 357, at 1088 (describing explanation as "a way to evaluate the basis of decision-making against broader normative constraints such as antidiscrimination or due process").

371 *See* Kroll et al., *supra* note 18, at 660 (describing how "transparency alone does little to explain either why any particular decision was made or how fairly the system operates").

372 *See id.* at 657-60; Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 4 (2018) (describing different types of gaming).

373 *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 53-65 (2016); Kroll et al., *supra* note 18, at 658 n.79.

gain a strategic advantage in submission.[374] Such gaming would thwart the objective of technology-assisted review — the efficient and reliable identification of submissions that are unlikely to benefit from further evaluation.

C. Imagining Potential Future Implementation

In a hypothetical future implementation, a machine learning system might be able to assign a score to a submission by analyzing substantive features associated with higher quality articles and identifying patterns in submissions using NLP techniques.[375] Based on the patterns, the system could assign a score to indicate the predicted strength of a submission, helping editors decide whether it should receive further evaluation. For example, one group of researchers has used machine learning techniques that include "domain-oriented features" to improve the automated evaluation of quality of medical articles on Wikipedia.[376]

A major problem with this imagined future application is the inability of artificial intelligence to carry out abstract analysis.[377] For example, in filtering spam messages, the goal is typically efficiency rather than accuracy.[378] By contrast, the use of proxies for quality to score submissions may result in false positives (where weaker submissions will be scored highly) and false negatives (where stronger articles will receive low scores).[379] These errors could result in substantial harm to authors and readers of scholarly literature.[380] Although the sheer number of law reviews and the availability of public databases for publications limit the risk that the public will be deprived of a paper,

---

374 *But see* Citron & Pasquale, *supra* note 327, at 11 (describing the difficulty of reverse engineering credit scores).

375 *See* Brian S. Haney, *Applied Natural Language Processing for Law Practice*, 2020 B.C. INTELL. PROP. & TECH. F. 1, 2.

376 VITTORIA COZZA, MARINELLA PETROCCHI & ANGELO SPOGNARDI, A MATTER OF WORDS: NLP FOR QUALITY EVALUATION OF WIKIPEDIA MEDICAL ARTICLES, *in* INTERNATIONAL CONFERENCE ON WEB ENGINEERING 448-56 (2016) (using machine learning techniques that consider "domain-relevant features" to automatically evaluate the quality of medical articles on Wikipedia). *But see* Checco et al., *supra* note 40, at 4 (observing that "assessing the quality of complex documents by automated means is still a challenging problem").

377 *See* Surden, *Machine Learning*, *supra* note 11, at 97-100 (observing that "many complicated problems . . . may not be amenable to such a heuristic-based technique").

378 *See id.* at 98.

379 *See id.* at 99-100.

380 *See id.*

readers also rely on heuristics, including law review placement, to determine which pieces to read and cite.[381]

Machine learning techniques used to identify substantive considerations in prior published articles are unlikely to provide accurate scores for submissions that are novel.[382] Where the number of examples provided in a given area are insufficient, such as in a nascent area of legal study, machine learning may fail to identify patterns that are reliable predictors of submissions that should receive a high score.[383] Partially automated screening tools also might not consider the improving quality of submissions in an emerging area over time.[384] Consequently, predicting the substantive elements that a strong submission should contain to receive a high score in a consistent and fair manner will be extremely difficult, if not impossible.

CONCLUSION

Artificial intelligence has the potential to increase efficiency while minimizing bias in the law review submissions process, but it may cause significant harm. At the present, editors can benefit from technology that simplifies preemption checking, detects plagiarism, evaluates compliance with technical requirements, and formats citations. By partially automating these tasks, artificial intelligence can provide editors the time they need to review submissions that are likely to benefit from additional consideration. Partially automated technology could also provide a means for evaluating selection decisions, revealing partiality in human reviewers.

Notwithstanding these potential benefits, using artificial intelligence in the submissions process involves substantial risks. A partially automated system may codify bias into the selection process if the data used in training the system does not reflect the breadth of submissions in different areas or if human biases are incorporated into the system. Some of these adverse outcomes can be addressed through attentive design, measured transparency, and regular audits. Despite its shortcomings, using artificial intelligence in the law review submissions

381 *See* Heifetz, *supra* note 7, at 632 (describing how the selection of articles by editors can "shape the professional literature for consumption by academics, judges, and practicing attorneys").

382 *See* Lehr & Ohm, *supra* note 14, at 680; Surden, *Machine Learning*, *supra* note 11, at 105.

383 *See* Shapiro & Pearse, *supra* note 254, at 1507; Shapiro, *Legal Scholars*, *supra* note 254, at 413.

384 *See* Checco et al., *supra* note 40, at 9.

process may still be worth considering in limited circumstances. With circumscribed application and careful oversight, technology-assisted review offers the potential to bring about an improved submissions experience.