Lawyer and Judicial Competency in the Era of Artificial Intelligence: Ethical Requirements for Documenting Datasets and Machine Learning Models

MARK L. SHOPE*

ABSTRACT

Judges and lawyers have the duty of technology competence, which includes competence in artificial intelligence technologies ("AI"). So not only must lawyers advise their clients on new legal, regulatory, ethical, and human rights challenges associated with AI, they increasingly need to evaluate the ethical implications of including AI technology tools in their own legal practice. Similarly, judge competence consists of, among other things, knowledge and skill of technology relevant to service as a judicial officer, which includes AI. After describing how AI implicates ethical issues for lawyers and judges and the requirement for lawyers and judges to have technical competency in the AI tools they use, this article argues for the requirement to use one or both of the following human interpretable AI disclosure forms when lawyers and judges are using AI tools: Dataset Disclosure Form or Model Disclosure Form.

TABLE OF CONTENTS

INTRODUCTION 192

I. LAWYER AND JUDGE COMPETENCE 193

II. PARTICULAR COMPETENCE FOR LAWYERS 194

A. RELEVANT RULES 196

III. PARTICULAR COMPETENCE FOR JUDGES 201

* Assistant Professor of Law, National Yang Ming Chiao Tung University School of Law (markshope@-nycu.edu.tw). This article is based on the author's presentation at The 2nd International Academic Conference on Artificial Intelligence and Law hosted by the International Artificial Intelligence and Law Research Foundation on December 26-27, 2019, in Taipei and Kaohsiung, Taiwan. The main topic of the conference was "The Collision of the Legal Profession and Technology in the Modern Era: International Inquiry into AI and Law."

A. RELEVANT RULES 202

IV. MACHINE LEARNING LIFECYCLE 204

V. PROBLEMS WITH ARTIFICIAL INTELLIGENCE 206

VI. DATASET DISCLOSURE FORMS AND MODEL DISCLOSURE FORMS 210

VII. DATASET DISCLOSURE FORMS. 212

VIII. MODEL DISCLOSURE FORMS 218

CONCLUSION. 222

INTRODUCTION

Judges and lawyers have the duty of technology competence, which includes competence in artificial intelligence ("AI").[1] In practice, not only must lawyers advise their clients on new legal, regulatory, ethical, and human rights challenges associated with AI, but they increasingly must also evaluate the ethical implications of including AI technology tools in their own legal practice.[2] Similarly, judge competence consists of, among other things, knowledge and skill of technology relevant to service as a judicial officer, which includes AI.[3] After describing how AI implicates ethical issues for lawyers and judges, and the requirement for lawyers and judges to have technical competency in the AI tools they use, this article argues for the use of dataset disclosure forms for datasets ("Dataset Disclosure Forms") and model disclosure forms for models ("Model Disclosure Forms") when lawyers and judges are using AI tools in their professional capacity. Currently, there is no standardized process for documenting datasets and models in AI.[4] If lawyers and judges are using AI tools, relevant state

1. *See* Agnieszka McPeak, *Disruptive Technology and the Ethical Lawyer*, 50 U. TOL. L. REV. 457, 457–58 (2019) (noting that the "duty of technological competence has come about at a time when innovation, often fueled by artificial intelligence, has produced new legal technology, . . . [which] require lawyers to adapt the very tools of their trade in order to stay competent").

2. *See* Catherine Nunez, *Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions*, 20 TUL. J. TECH. & INTELL. PROP. 189, 195 (2017) (discussing various theories of a lawyer's role in legal ethics).

3. *See* Michael Thomas Murphy, *Just and Speedy: On Civil Discovery Sanctions for Luddite Lawyers*, 25 GEO. MASON L. REV. 36, 52 (2017) (observing the increases in technological competency of judges).

4. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III & Kate Crawford, *Datasheets for Datasets* 2 (Working Paper, No. 1803.09010v7, 2020), https://arxiv.org/pdf/1803.09010.pdf [https://perma.cc/M4EW-ARXD]. The author would like to extend a special

professional conduct rules for lawyers and judges create the obligation to understand these tools.[5] However, if AI tool providers do not divulge the inner working of those AI tools, then how can lawyers and judges truly understand these tools? To begin to answer this critical inquiry, this article argues for the use of Dataset Disclosure Forms and Model Disclosure Forms to fulfill lawyers' and judges' ethical requirement to understand the tools they are using. A Dataset Disclosure Form or a Model Disclosure Form may help stakeholders make informed decisions about the use of a particular AI tool and its flaws; thus, such forms should be included in relevant court procedures.

I. LAWYER AND JUDGE COMPETENCE

The ABA Model Rules of Professional Conduct (the "Model Rules") include the duty of technology competence in Rule 1.1, Comment 8.[6] Since AI is one kind of technology, this technology competence would include competence in AI technologies relevant to the practitioner's legal practice.[7] Lawyers must evaluate how and to what extent they are including AI technology tools in their own legal practice.[8] For example, Model Rules 1.4 (Communications) or 1.5 (Fees) could be triggered if a lawyer thinks using a particular AI solution would be useful for a client's legal needs and when the lawyer is discussing fees. Rule 1.6 (Confidentiality of Information) is triggered when a lawyer is using an AI solution that is not in-house, possibly exposing confidential information to a third-party (for example, when electronic document review is conducted off-site by a third-party). Communicating with the court and opposing counsel regarding electronic discovery could implicate Rules 3.3 (Candor to the Tribunal) and 3.4 (Fairness to Opposing Party & Counsel). Rule 5.1 (Responsibilities of a Partner or Supervisory Lawyer) and Rule 5.3 (Responsibilities Regarding Nonlawyer Assistants) could be triggered when a subordinate is tasked with deciding which particular AI tool to use and further while implementing those tools. Rule 5.5

---

thanks to the authors of this article, including Timnit Gebru, who approved the use of a portion of their groundbreaking paper to develop the Dataset Disclosure Forms and Model Disclosure Forms found below.

5. *See* Marla N. Greenstein, *AI and a Judge's Ethical Obligations*, 59 Judges' J. (forthcoming Winter 2020) (noting that for judges to comply "with their ethical responsibilities while using AI or interpreting its proper use, [they] must first ensure that they understand the AI application involved"). This article focuses on a practical implementation of relevant model rules by discussing those rules as they are manifest in Indiana.

6. *See* MODEL RULES OF PROF'L CONDUCT R. 1.1 cmt. 8 (2020). [hereinafter MODEL RULES].

7. Ed Walters, *The Model Rules of Autonomous Conduct: Ethical Responsibilities of Lawyers and Artificial Intelligence*, 35 GA. ST. U. L. REV. 1073, 1076 (2019) (noting that "as the quality of work product created by lawyers augmented with AI surpasses the work created without AI, it is clear that lawyers will soon have a professional responsibility to employ new techniques").

8. *See* Katherine Medianik, *Artificially Intelligent Lawyers: Updating the Model Rules of Professional Conduct in Accordance with the New Technological Era*, 39 CARDOZO L. REV. 1497, 1501–02 (2018) (proposing, given the growing technology in the legal field, "(1) the addition of continuing legal education (CLE) requirements on 'Legal Technology'; (2) the addition of the term 'nonlawyer assistant' to the terminology section of the Model Rules [of Professional Conduct]; and (3) the addition of several comments that incorporate AI technology and account for technological advancement").

(Unauthorized Practice of Law) may be implicated when a non-lawyer relies on an AI tool that renders what would be considered legal advice. Rule 5.7 (Responsibilities Regarding Law-Related Services) may be triggered when the lawyer provides an AI-utilized service for a client which may not be considered traditional legal services. Screening lawyer candidates using an algorithmic recruiting solution which is biased may be considered discrimination, implicating Rule 8.4 (Misconduct).

Judges must also understand AI technologies and how these technologies affect the judge's conduct and docket.[9] For example, Rule 2.2 (Impartiality and Fairness) could be triggered when a judge is using an AI solution that is considered impartial or unfair and sufficiently influences his or her judgment. Rule 2.3 (Bias, Prejudice, and Harassment) could be triggered when a judge is using an AI solution that has bias in the algorithm or training data. Rule 2.4 (External Influences on Judicial Conduct) could be triggered when a judge is using an AI solution based on external pressure to use the technology. Rule 2.5 (Competence, Diligence, and Cooperation) expresses the general requirement for judges to know the benefits and risks associated with the technology relevant to service as a judicial officer. Rule 2.13 (Hiring and Administrative Appointments) could be triggered when screening clerks or other administrative candidates using an algorithmic recruiting solution which is biased because this may be considered discrimination.

II. PARTICULAR COMPETENCE FOR LAWYERS

This article focuses on a practical implementation of the ABA Model Rules by discussing the Indiana Rules of Professional Conduct (the "Indiana Lawyer Rules").[10] Rule 1.1 of the Indiana Lawyer Rules, which adopts the language of the Model Rules, requires a lawyer to provide "competent representation" in rendering legal services to a client.[11] "Competent representation" is described as "the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation."[12] In particular, Comment 6 of Rule 1.1 of the Indiana Lawyer Rules states that "a lawyer should keep abreast of changes in the law and its practice, *including the benefits and risks associated with the technology relevant to the lawyer's practice*, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject."[13] This comment establishes the principle of technological competence

9. *See* MODEL CODE OF JUDICIAL CONDUCT (Am. Bar Ass'n 2010).

10. IND. RULES OF PROF'L CONDUCT (Ind. Rules of Court July 03, 2019). This Article focuses on the ethical rules as they are manifest in Indiana, which largely adopted the ABA Model Rules of Professional Conduct in 1986. *See* Jurisdictional Rules Comparison Charts, ABA, https://www.americanbar.org/groups/professional_responsibility/policy/rule_charts/ [https://perma.cc/6NLM-P8WF].

11. IND. RULES OF PROF'L CONDUCT R. 1.1.

12. IND. RULES OF PROF'L CONDUCT R. 1.1.

13. IND. RULES OF PROF'L CONDUCT R. 1.1 cmt. 6 (emphasis added).

in the lawyer's profession.[14] Competence is the key principle, but it does not (always) require attorneys to be experts.[15] Competence includes knowing the limits of the lawyer's own understanding (and understanding the evolving nature of the technology), seeking assistance (in accordance with his or her ethical obligations), and demanding more cooperation from outside service providers. Lawyers must also understand what an AI tool can and cannot do and be able to know the extent to which an AI tool is effective and accomplishes its objective, for example, when an attorney uses technology assisted review to find responsive, privileged, or confidential documents.[16]

In some situations, the lawyer may be required to have a higher level of competence depending on the situation.[17] The comments to the Indiana Lawyer Rules list factors relevant in determining whether a lawyer employs the requisite knowledge and skill in a particular matter.[18] These factors take into consideration "the relative complexity and specialized nature of the matter, the lawyer's general experience, the lawyer's training and experience in the field in question, [and] the preparation and study the lawyer is able to give the matter[.]"[19] Importantly, the comment also states that "[e]xpertise in a particular field of law may be required in some circumstances."[20] Furthermore, "[c]ompetent handling of a particular matter includes . . . *use of methods and procedures* meeting the standards of competent practitioners."[21] This comment suggests that in some situations the lawyer, when providing AI technology related advice to a client, may need *expertise in* AI technology to employ the requisite knowledge and skill to meet the requirements of the Indiana Lawyer Rules. Furthermore, utilizing "methods and procedures" necessary to meet the standards of a competent practitioner arguably includes competency in AI technology if the lawyer decides to use AI tools.[22]

---

14. *See* Daniel N. Kluttz & Deirdre K. Mulligan, *Automated Decision Support Technologies and the Legal Profession*, 34 BERKELEY TECH. L.J. 853, 868 (2019) (noting that "[i]n 2012, the legal profession began the process of establishing a legal duty of technological competence on lawyers") (emphasis omitted).

15. Ronald J. Hedges & Amy Walker Wagner, *Competence with Electronically Stored Information: What Does It Currently Mean in the Context of Litigation and How Can Attorneys Achieve It?*, The American Law Institute Continuing Legal Education, ALI-CLE Course Materials, SX029 ALI-CLE 1107 (2016) (describing that the matter dictates the level of competence).

16. *See* James A. Sherer & Ed Walters, *Practical Magic: Law's Hands-on AI Revolution*, 44 LAW PRAC. 32, 36 (2018) (noting that "attorneys who incorporate [AI or near-AI] use into their decision making must understand the methods and the design of those AI tools in order to understand that they, too, are not working by magic but by design according to some original direction based on certain data sets, assumptions and a general sense of application").

17. Judith Welch Wegner, *Contemplating Competence: Three Meditations*, 50 VAL. U. L. REV. 675, 702 (2016) (discussing level of competencies according to field of practice).

18. IND. RULES OF PROF'L CONDUCT R. 1.1 cmt. 1.

19. IND. RULES OF PROF'L CONDUCT R. 1.1 cmt. 1.

20. IND. RULES OF PROF'L CONDUCT R. 1.1 cmt. 1.

21. IND. RULES OF PROF'L CONDUCT R. 1.1 cmt. 5.

22. Drew Simshaw, *Ethical Issues in Robo-Lawyering: The Need for Guidance on Developing and Using Artificial Intelligence in the Practice of Law*, 70 HASTINGS L.J. 173, 211 (2018) (noting that "[c]ompetence in the era of AI should require a lawyer to either be involved in the design of the AI systems they are using, or at

Lawyers must therefore understand AI technologies and associated processes and in some cases, technical expertise is necessary. In short, traditional legal prowess is not enough to evaluate or use AI technologies. Specific knowledge is required and needs to be continually updated.[23]

This Article will discuss specific situations in which the Indiana Lawyer Rules could be triggered when using AI technology.

A. RELEVANT RULES

Rule 1.4 (Communications)[24] could be triggered if a lawyer thinks using an AI solution would be useful for a client's legal needs.[25] If the lawyer thinks that using a particular AI tool will help reach the client's objectives, the lawyer has a duty to keep the client reasonably informed about such matters.[26] In addition, the rules may require the lawyer to explain a particular AI tool so that the client can make an informed decision regarding the use of such technology.[27] This requires the lawyer to be knowledgeable about the AI tool before consulting with the client.[28]

Machine learning is becoming more accurate; as a result, algorithms could eventually be used to predict settlement value. Details of a settlement must be communicated to clients.[29] With regard to fees, Rule 1.5 (Fees) requires that if a lawyer thinks using a particular AI solution would be useful for a client's legal needs, the lawyer must communicate these technology expenses to the client.[30]

---

the very least, to understand—with the help of an expert, if needed—certain underlying characteristics that affect (1) the AI's bias . . . ; (2) AI's limits . . . ; and (3) AI's confidentiality concerns").

23. John Flood & Lachlan Robb, *Professions and Expertise: How Machine Learning and Blockchain Are Redesigning the Landscape of Professional Knowledge and Organization*, 73 U. MIAMI L. REV. 443, 448 (2019) (noting that legal professionals "embody expertise and knowledge . . . by the content of their education and training").

24. IND. RULES OF PROF'L CONDUCT R. 1.4 (noting that a lawyer must "reasonably consult with the client about the means by which the client's objectives are to be accomplished" and "explain a matter to the extent reasonably necessary to permit the client to make informed decisions regarding the representation").

25. IND. RULES OF PROF'L CONDUCT R. 1.4 (noting that a client "should have sufficient information to participate intelligently in decisions concerning the objectives of the representation and the means by which they are to be pursued, to the extent the client is willing and able to do so").

26. IND. RULES OF PROF'L CONDUCT R. 1.4 cmt. 3 (noting that a lawyer must "reasonably consult with the client about the means to be used to accomplish the client's objectives . . . [, which in some situations] will require consultation prior to taking action").

27. IND. RULES OF PROF'L CONDUCT R. 1.4 cmt. 6 (stating that "the information to be provided is that appropriate for a client who is a comprehending and responsible adult").

28. Kurt Watkins & Rachel E. Simon, *AI and the Young Attorney: What to Prepare for and How to Prepare*, 11 LANDSLIDE 22, 26 (2019) ("With intimate knowledge of the available AI tool, an attorney will be able to select the best tool and the best way to use the tool for each client and matter . . . [W]ith the appropriate AI tools, the attorney will develop greater knowledge of fitting AI work product to client needs").

29. IND. RULES OF PROF'L CONDUCT R. 1.4 cmt. 2 (noting that "a lawyer who receives from opposing counsel an offer of settlement in a civil controversy . . . must promptly inform the client of its substance unless the client has previously indicated that the proposal will be acceptable or unacceptable or has authorized the lawyer to accept or to reject the offer").

30. IND. RULES OF PROF'L CONDUCT R.1.5(b) (noting that the "basis or rate of the fee and expenses for which the client will be responsible shall be communicated to the client, preferably in writing, before or within a reasonable time after commencing the representation").

Using due diligence tools, prediction technology, legal analytics tools, document automation tools, and intellectual property tools will certainly add expenses to client bills.[31] But whether these technologies will reduce or increase total bills may depend on whether the particular tool reduces attorney fees.[32] Either way, an increase or decrease in total cost needs to be reasonably communicated to clients.[33]

Rule 1.6 (Confidentiality of Information) could be triggered when a lawyer is using an AI solution that is not in-house (for example, when electronic document review is conducted off-site by a third-party).[34] Problems could also arise when lawyers are using AI to assist in production during the discovery stage.[35] Machine learning is not perfect and attorney-client privileged communications or attorney work product could sneak into a production.[36]

In the former situation (when electronic document review is conducted off-site by a third-party), there is usually an exception to this rule for certain authorized disclosures "when appropriate in carrying out the representation."[37] Lawyers often use this provision (and disclose this in engagement letters) to engage legal service providers who are acting under the direction of the lawyer.[38] In the latter situation (when attorney-client privileged communications or attorney work product sneak into a production), the lawyer did not properly safeguard the information and should take steps to remedy the situation.[39]

---

31. Sean Semmler & Zeeve Rose, *Artificial Intelligence: Application Today and Implications Tomorrow*, 16 DUKE L. & TECH. REV. 85, 90 (2017).

32. *Id.* at 89–90.

33. IND. RULES OF PROF'L CONDUCT R. 1.5(b) cmt. 2 (noting that "an understanding as to fees and expenses must be promptly established[, by, for example, furnishing] the client with at least a simple memorandum or copy of the lawyer's customary fee arrangements that states the general nature of the legal services to be provided, the basis, rate or total amount of the fee and whether and to what extent the client will be responsible for any costs, expenses or disbursements in the course of the representation").

34. IND. RULES OF PROF'L CONDUCT R. 1.6(a) ("A lawyer shall not reveal information relating to representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation or the disclosure is [otherwise permitted]").

35. Robert Keeling, Nathaniel Huber-Fliflet, Jianping Zhang & Rishi P. Chhatwal, *Separating the Privileged Wheat from the Chaff - Using Text Analytics and Machine Learning to Protect Attorney-Client Privilege*, 25 RICH. J.L. & TECH. 2, 27 (2019) (noting that "[p]redictive modeling has proven to be very effective at identifying relevant documents, but there is a widely held belief in the legal community that it is incapable of mimicking the nuanced analysis required for privilege decisions").

36. *Id.* at 12 (noting that "[i]nadvertent disclosure is by far the most common method of waiver" of the attorney-client privilege).

37. IND. RULES OF PROF'L CONDUCT R. 1.6 cmt. 5 (noting that except for certain circumstances "a lawyer is impliedly authorized to make disclosures about a client when appropriate in carrying out the representation").

38. *See generally* Roland L. Trope & Sarah Jane Hughes, *Red Skies in the Morning—Professional Ethics at the Dawn of Cloud Computing*, 38 WM. MITCHELL L. REV. 111 (2011) (discussing professional ethics relating to cloud computing).

39. IND. RULES OF PROF'L CONDUCT R. 1.6 cmt. 16 (noting that "[a] lawyer must act competently to safeguard information relating to the representation of a client against inadvertent or unauthorized disclosure by the lawyer or other persons who are participating in the representation of the client or who are subject to the lawyer's supervision").

Rule 3.3 (Candor to the Tribunal)[40] and Rule 3.4 (Fairness to Opposing Party & Counsel)[41] could be triggered when communicating with the court and opposing counsel regarding electronic discovery. A lawyer must not "unlawfully obstruct another party's access to evidence or unlawfully alter, destroy or conceal a document or other material having potential evidentiary value."[42] In order to fulfill these requirements, the lawyer must be knowledgeable of the relevant technologies (often powered by AI) available to locate, isolate, and produce that evidence.[43] If a lawyer is using an AI tool, the lawyer must be familiar with the relevant technologies to demonstrate full candor toward the tribunal and not "knowingly . . . make a false statement of fact or law to a tribunal or fail to correct a false statement of material fact or law previously made to the tribunal by the lawyer."[44]

Lawyers also have an obligation to protect a tribunal "against criminal or fraudulent conduct that undermines the integrity of the adjudicative process, such as . . . unlawfully destroying or concealing documents or other evidence or failing to disclose information to the tribunal when required by law to do so."[45] The lawyer must "take reasonable remedial measures, including disclosure if necessary, whenever the lawyer knows that a person, including the lawyer's client, intends to engage, is engaging or has engaged in criminal or fraudulent conduct related to the proceeding."[46] If the lawyer is using an AI tool to find relevant documents or evidence and the lawyer does not know the full extent to which the AI tool is capturing relevant documents or evidence, the AI tool (and by extension the lawyer) could be contributing to concealing those relevant documents and evidence.[47] The idea behind this is that "the adversary system contemplates that the evidence in a case is to be marshaled competitively by the contending parties. Fair

40. IND. RULES OF PROF'L CONDUCT R. 3.3(a) (noting that a "lawyer shall not knowingly . . . make a false statement of fact or law to a tribunal or fail to correct a false statement of material fact or law previously made to the tribunal by the lawyer . . . [, or] offer evidence that the lawyer knows to be false").

41. IND. RULES OF PROF'L CONDUCT R. 3.4 ("A lawyer shall not: (a) unlawfully obstruct another party's access to evidence or unlawfully alter, destroy or conceal a document or other material having potential evidentiary value . . . ; (b) falsify evidence, counsel or assist a witness to testify falsely, or offer an inducement to a witness that is prohibited by law; (c) knowingly disobey an obligation under the rules of a tribunal except for an open refusal based on an assertion that no valid obligation exists; [or] (d) in pretrial procedure, make a frivolous discovery request or fail to make reasonably diligent effort to comply with a legally proper discovery request by an opposing party").

42. IND. RULES OF PROF'L CONDUCT R. 3.4(a).

43. *See* Judith L. Maute, *Facing 21st Century Realities*, 32 MISS. C. L. REV. 345, 369 (2013) (noting that "if a lawyer cannot master the technology suitable for that lawyer's practice, the lawyer should either hire tech-savvy lawyers tasked with responsibility to keep current, or hire an outside technology consultant who understands the practice of law and associated ethical constraints").

44. IND. RULES OF PROF'L CONDUCT R. 3.3(a)(1).

45. IND. RULES OF PROF'L CONDUCT R. 3.3 cmt. 12.

46. IND. RULES OF PROF'L CONDUCT R. 3.3 cmt. 12.

47. *The Case for Cooperation*, 10 SEDONA CONF. J. 339, 344 (2009) (noting that "refusing to 'aid' opposing counsel in designing an appropriate search protocol that the party holding the data knows will produce responsive documents could be tantamount to concealing relevant evidence").

competition in the adversary system is secured by prohibitions against destruction or concealment of evidence, . . . obstructive tactics in discovery procedure, and the like."[48] Having access to all relevant information is used to establish claims and defenses, so the right of a party to obtain evidence is an essential procedural right that "can be frustrated if relevant material is altered, concealed or destroyed."[49]

Rule 5.1 (Responsibilities of a Partner or Supervisory Lawyer) could be triggered when a subordinate is tasked with deciding which AI tool to use and further while implementing those tools.[50] When utilizing AI technology, a supervising lawyer must have knowledge of this technology to fulfill the supervising lawyer's managerial responsibilities for lawyers under supervision and for nonlawyer assistants under supervision that are utilizing these same technologies.[51] Methods vary to comply with this rule.[52] Sometimes informal supervision and periodic reviews of compliance will suffice.[53] In other situations, these informal measures may not be enough and more elaborate measures may be required.[54]

This rule also covers nonlawyers employed by a law firm.[55] In this situation, the supervising lawyer must "make reasonable efforts to ensure that the [nonlawyer's] conduct is compatible with the professional obligations of the lawyer."[56] These efforts include instruction and supervision regarding all aspects of the lawyers ethical requirements, including taking responsibility for the work product of the nonlawyer.[57] This would likely include the requirement for the supervising lawyer to have sufficient knowledge about any AI tools that are in use by the nonlawyer, so that the supervising lawyer could give adequate instruction and supervision over those AI tools. Lawyers in supervisory roles must also take into account that the nonlawyers they supervise probably do not have formal legal training, so adequate instruction and supervision relating to the AI tool and its ethical considerations is particularly important.[58]

---

48. IND. RULES OF PROF'L CONDUCT R. 3.4 cmt. 1.

49. IND. RULES OF PROF'L CONDUCT R. 3.4 cmt. 2.

50. IND. RULES OF PROF'L CONDUCT R. 5.1(a) (noting that lawyers with managerial authority "shall make reasonable efforts to ensure that the firm has in effect measures giving reasonable assurance that all lawyers in the firm conform to the Rules of Professional Conduct").

51. For a novel approach to expressing AI related issues in the model rules, *see* Medianik, *supra* note 8, at 1529 (proposing language to the comment to the ABA model rule 5.3 that "adds AI technology within the scope of the traditional meaning of nonlawyer assistant and provides instructions to supervising lawyers that they must supervise AI technology as they would human nonlawyer assistants").

52. Richard B. Polony & Brendan J. McCartney, *Is It Safe? Ethical Implications of Connectivity*, 21 FIDELITY L.J. 37, 53 (2015) (noting that methods "depend on the size of the firm, structure, and practice").

53. IND. RULES OF PROF'L CONDUCT R. 5.1 cmt. 3.

54. IND. RULES OF PROF'L CONDUCT R. 5.1 cmt. 3.

55. *See* Douglas R. Richmond, *Watching Over, Watching Out: Lawyers' Responsibilities for Nonlawyer Assistants*, 61 U. KAN. L. REV. 441, 445 (2012) (noting that "[l]awyers' duties to supervise nonlawyer assistants . . . closely parallel their duties to supervise fellow lawyers").

56. IND. RULES OF PROF'L CONDUCT R. 5.3(b).

57. IND. RULES OF PROF'L CONDUCT R. 5.3 cmt. 1.

58. *See* IND. RULES OF PROF'L CONDUCT R. 5.3 cmt. 1. "[L]awyers with managerial authority . . . [must] make reasonable efforts to establish internal policies and procedures designed to provide reasonable assurance

Rule 5.5 (Unauthorized Practice of Law)[59] could be triggered when a nonlawyer relies on an AI tool that renders what would be considered legal advice.[60] Of course, whether an AI tool constitutes the "practice of law" for unauthorized practice of law purposes is the threshold question.[61] This question is outside the scope of this article.[62] If use of an AI tool is considered legal practice, nonlawyers providing an AI tool to advise third parties may be engaging in the practice of law and could be subject to criminal penalties.[63] In addition, lawyers may unknowingly be assisting in the unauthorized practice of law if they are not sufficiently supervising this process.[64]

Rule 5.7 (Responsibilities Regarding Law-Related Services)[65] could be triggered when the lawyer provides an AI-utilized service for a client which may not be considered traditional legal services.[66] Serving the client often includes what would be considered law-related services.[67] These law-related services could include, among other things, "providing title insurance, financial planning,

---

that nonlawyers in the firm will act in a way compatible with the Rules of Professional Conduct." IND. RULES OF PROF'L CONDUCT R. 5.3 cmt. 2.

59. IND. RULES OF PROF'L CONDUCT R. 5.5 ("A lawyer shall not practice law in a jurisdiction in violation of the regulation of the legal profession in that jurisdiction, or assist another in doing so.").

60. *See* IND. RULES OF PROF'L CONDUCT R. 5.5 cmt. 1 & 2 (noting that a "lawyer may practice law only in a jurisdiction in which the lawyer is authorized to practice").

61. Restatement (Third) of the Law Governing Lawyers § 4 (2000) ("To some, the expression 'unauthorized practice of law' by a nonlawyer is incongruous, because it can be taken to imply that nonlawyers may engage in some aspects of law practice, but not others. The phrase has gained near-universal usage in the courts, ethics-committee opinions, and scholarly writing, and it is well understood not to imply any necessary area of permissible practice by a nonlawyer. Moreover, a nonlawyer undoubtedly may engage in some limited forms of law practice, such as self-representation in a civil or criminal matter. . . . It thus would not be accurate for the black letter to state flatly that a nonlawyer may not engage in law practice. . . . A nonlawyer who impermissibly engages in the practice of law may be subject to several sanctions, including injunction, contempt, and conviction for crime.").

62. *See generally* Thomas E. Spahn, Partner, McGuireWoods LLP, Legal Ethics and Artificial Intelligence: Final Frontier or Today's Reality? Seminar, The American Law Institute Continuing Legal Education, ALI-CLE Course Materials (February 22, 2018).

63. Thomas E. Spahn, *Is Your Artificial Intelligence Guilty of the Unauthorized Practice of Law?*, 24 RICH. J.L. & TECH. 2, 2 (2018) (noting that "non-lawyers relying on AI to advise third parties may be committing the criminal unauthorized practice of law, and lawyers insufficiently involved in such a process may be guilty of assisting in such unauthorized practice of law").

64. *Id.* at 19 (noting that "lawyers involve themselves with a non-lawyer's use of artificial intelligence, they may also face allegations that they are assisting in the unauthorized practice of law by not adequately supervising and approving such non-lawyer efforts").

65. IND. RULES OF PROF'L CONDUCT R. 5.7 ("(a) A lawyer shall be subject to the Rules of Professional Conduct with respect to the provision of law-related services [under some circumstances]. . . . (b) The term 'law-related services' denotes services that might reasonably be performed in conjunction with and in substance are related to the provision of legal services, and that are not prohibited as unauthorized practice of law when provided by a non-lawyer.").

66. *See* IND. RULES OF PROF'L CONDUCT R. 5.7 cmt. 1 ("When a lawyer performs law-related services. . . , there exists the potential for ethical problems. Principal among these is the possibility that the person for whom the law-related services are performed fails to understand that the services may not carry with them the protections normally afforded as part of the client-lawyer relationship.").

67. IND. RULES OF PROF'L CONDUCT R. 5.7 cmt. 9.

accounting, real estate counseling, legislative lobbying, economic analysis, social work, psychological counseling, tax preparation, and medical or environmental consulting."[68] Law-related services could also include providing AI or other technology services which serve the client in nonlegal ways.[69] For example, this could include fintech title insurance, using AI to improve financial planning, AI-enabled systems for accounting, AI-driven data points for real estate transactions, etc.

Rule 8.4 (Misconduct) could be triggered in a number of ways.[70] For example, screening lawyer candidates using a biased algorithmic recruiting solution[71] may be considered discrimination as well as using an AI for targeted online advertising for lawyers.[72] The rule could be triggered if a lawyer is using an AI tool that is considered to have bias in either its algorithm or training data, and the lawyer is using the tool in a case.

III. PARTICULAR COMPETENCE FOR JUDGES

Judges have a similar requirement for technology competence. Rule 2.5 of the *Indiana Code of Judicial Conduct* (the "Indiana Judge Rules"), which adopts the language of the *Model Code of Judicial Conduct*, states: "(A) A judge shall perform judicial and administrative duties competently, diligently, and promptly. (B) A judge shall cooperate with other judges and court officials in the administration of court business."[73] Comment 1 to this rule states that "[c]ompetence in the performance of judicial duties requires the legal knowledge, skill, thoroughness, and preparation reasonably necessary to perform a judge's responsibilities

68. IND. RULES OF PROF'L CONDUCT R. 5.7 cmt. 9.

69. Steven C. Bennett, *The Ethics of Electronic Discovery Recent Developments Demonstrate the Need to Establish "Best Practices,"* PRAC. LITIGATOR, Mar. 2006, at 45, 46 (noting that "some law firms have begun to offer their clients in-house e-discovery services, or are teaming with software designers to develop customized e-discovery software").

70. IND. RULES OF PROF'L CONDUCT R. 8.4 ("It is professional misconduct for a lawyer to: (a) violate or attempt to violate the Rules of Professional Conduct, knowingly assist or induce another to do so, or do so through the acts of another; . . . (c) engage in conduct involving dishonesty, fraud, deceit or misrepresentation; (d) engage in conduct that is prejudicial to the administration of justice; . . . or (g) engage in conduct, in a professional capacity, manifesting, by words or conduct, bias or prejudice based upon race, gender, religion, national origin, disability, sexual orientation, age, socioeconomic status, or similar factors.").

71. *See* McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, 71 ARK. L. REV. 529, 530 (2018) (noting that "[i]f employers wish to take advantage of the potential efficiency benefits of using artificial intelligence in hiring, they should use caution in selecting a program, encourage the use of responsible algorithms, and push for long term changes in the lack of racial and gender diversity in the technology industry").

72. Richard C. Klein, *The Advent of Social Networking: Addressing the Model Rules on Attorney Advertising and Solicitation*, 5 BIOTECHNOLOGY & PHARMACEUTICAL L. REV. 60, 63 (2012).

73. IND. CODE OF JUDICIAL CONDUCT R. 2.5 (adopted effective March 1, 1993, including amendments through May 16, 2019). This article focuses on the ethical rules as they are manifest in Indiana, which largely adopts the *ABA Model Code of Judicial Conduct*. *See* Charts Comparing Individual Jurisdictional Judicial Conduct Rules to *ABA Model Code of Judicial Conduct*, ABA, https://www.americanbar.org/groups/professional_responsibility/resources/judicial_ethics_regulation/aba_model_code_comparison/ [hhttps://perma.cc/25ZD-ZF36].

of judicial office, *including the benefits and risks associated with the technology relevant to service as a judicial officer*."[74] So not only must judges understand the new legal, regulatory, ethical, and human rights challenges associated with AI, they increasingly need to evaluate how they are (themselves or through the parties) including AI technology tools in their own docket.[75] This could include basic docket management and courtroom tools (like AI transcribing tools), risk assessment tools (in making decisions on sentencing, pretrial release, and parole), and understanding the science and law relating to electronically stored information and e-discovery.

This Article will discuss specific situations in which the Indiana Judge Rules could be triggered when using AI technology.

A. RELEVANT RULES

Rule 2.2 (Impartiality and Fairness) could be triggered when a judge is using or allowing a party to use an AI solution that is partial or unfair.[76] "[J]udges must understand the role that AI and machine learning play in the legal system itself."[77] For example, if one party is using an AI tool to find relevant documents or evidence and the judge (or the party) does not know the full extent to which the AI tool is capturing relevant documents or evidence, a judge may not be performing duties fairly and impartially and could actually be contributing to this unfairness and partiality.[78]

Rule 2.3 (Bias, Prejudice, and Harassment)[79] could be triggered when a judge is using or permitting the use of an AI solution that has bias in the algorithm or training data.[80] This bias can come from conduct, which can include using an AI tool that manifests bias or prejudice. For example, if using a risk assessment tool like the Correctional Offender Management Profiling for Alternative Sanctions, or COMPAS, manifests bias or prejudice, then by extension the judge could be

---

74. IND. CODE OF JUDICIAL CONDUCT R. 2.5 cmt. 1 (emphasis added).

75. *See* American Bar Association, House of Delegates, Resolution 112 (Aug. 12-13, 2019) (urging courts "to address the emerging ethical and legal issues related to the usage of artificial intelligence ('AI') in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI").

76. IND. CODE OF JUDICIAL CONDUCT R. 2.2 (noting that a "judge shall uphold and apply the law, and shall perform all duties of judicial office fairly and impartially").

77. Melissa Whitney, *How to improve technical expertise for judges in AI-related litigation*, BROOKINGS (Nov. 7, 2019), https://www.brookings.edu/research/how-to-improve-technical-expertise-for-judges-in-ai-related-litigation/ [https://perma.cc/ZX75-SDDP].

78. Greenstein, *supra* note 5, at 40 (noting that "[w]isdom will require the ability to use artificial intelligence to enhance integrity and impartiality, tempered by human judgment").

79. IND. CODE OF JUDICIAL CONDUCT R. 2.3 ("(A) A judge shall perform the duties of judicial office, including administrative duties, without bias or prejudice. (B) A judge shall not, in the performance of judicial duties, by words or conduct manifest bias or prejudice, or engage in harassment . . . (C) A judge shall require lawyers in proceedings before the court to refrain from manifesting bias or prejudice, or engaging in harassment. . . .").

80. IND. CODE OF JUDICIAL CONDUCT R. 2.3 cmt. 1 (noting that a "judge who manifests bias or prejudice in a proceeding impairs the fairness of the proceeding and brings the judiciary into disrepute").

engaging in such bias or prejudice.[81] This example underscores the need for judges to understand the technology that they are using or allowing to be used in their courtroom.[82]

Rule 2.4 (External Influences on Judicial Conduct)[83] could be triggered when a judge is using an AI solution that has a strong "popular" or "unpopular" sentiment among the public, and this sentiment sways the judge to use or not use a certain AI tool.[84] For example, the use of COMPAS has received criticism.[85] This criticism should not sway the judge; at most it could encourage the judge to probe the arguments for and against the use of COMPAS, including understanding any potential biases.[86]

Performing judicial and administrative duties competently, diligently, and promptly includes understanding the benefits and risks associated with the technology relevant to these duties.[87] Comment one illustrates this by stating that competence includes "the legal knowledge, skill, thoroughness, and preparation reasonably necessary to perform a judge's responsibilities of judicial office, including the benefits and risks associated with the *technology relevant to service as a judicial officer*."[88] Fortunately for a busy judge, this could include hiring those with the requisite expertise in relevant legal technology. Comment two of the Indiana Judge Rules demonstrates this by stating that "[a] judge should seek the necessary docket time, court staff, expertise, and resources to discharge all adjudicative and administrative responsibilities."[89] This introduces the possibility

81. *See* Taylor B. Schaefer, *The Ethical Implications of Artificial Intelligence in the Law*, 55 GONZ. L. REV. 221, 229 (2020) ("If an algorithm that a judge utilizes in making sentencing decisions produces racially-skewed results, the judge could potentially be in violation of [the Code of Judicial Conduct] by allowing impartiality to infiltrate the judicial process.").

82. *See* Lisa A. Hayes, *Technology in the Courts: More Questions Than Settled Legal Answers*, 42 HUM. RTS. 19, 22 (2017) (explaining that a judge "must do his or her best to become familiar with the underlying information technology, think about its sweeping implications, and understand how most of our society engages with the product or service").

83. IND. CODE OF JUDICIAL CONDUCT R. 2.4 (noting that a "judge shall not be swayed by public clamor or fear of criticism").

84. IND. CODE OF JUDICIAL CONDUCT R. 2.4 cmt. 1 ("An independent judiciary requires that judges decide cases according to the law and facts, without regard to whether particular laws or litigants are popular or unpopular with the public, the media, government officials, or the judge's friends or family.").

85. *See State v. Loomis*, 881 N.W.2d 749, 754 (Wis. 2016) (explaining circuit courts should use caution when using COMPAS because it only identifies high risk groups, not individuals); *see also* Recent Case, *Criminal Law, State v. Loomis, Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*, 130 HARV. L. REV. 1530, 1532 (2017).

86. Katherine Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, 18 N.C.J.L. & TECH. ONLINE 75, 103–04 (2016) (discussing training for judges who want to use COMPAS).

87. IND. CODE OF JUDICIAL CONDUCT R. 2.5 ("(A) A judge shall perform judicial and administrative duties competently, diligently, and promptly. (B) A judge shall cooperate with other judges and court officials in the administration of court business.").

88. IND. CODE OF JUDICIAL CONDUCT R. 2.5 cmt. 1 (emphasis added).

89. IND. CODE OF JUDICIAL CONDUCT R. 2.5 cmt. 2.

of judges being able to hire a full-time expert in AI technologies to assist with various court challenges relating to emerging AI technologies.

Rule 2.13 (Hiring and Administrative Appointments)[90] could be triggered when screening administrative and clerk candidates using a biased algorithmic recruiting solution that may be considered discrimination.[91]

IV. MACHINE LEARNING LIFECYCLE

Before discussing problems with machine learning, this article will briefly describe the machine learning lifecycle. There could be a number of ways to describe the structure of the machine learning lifecycle based on the specific project. This article generally focuses on the Suresh & Guttag (2020) lifecycle describing five steps: data collection, data preparation, model development, model evaluation and postprocessing, and model deployment.[92] The careful reader will note during each step how data is manipulated, and how this manipulation could invite unreliable results. Judges and lawyers should understand this process because it provides context to the potential sources of bias that will be discussed. It also provides context to the Dataset Disclosure Form and Model Disclosure Form described below.

Data Collection: Data is collected from the world and involves gathering a population of records and assigning features and labels to use on the records.[93] Together, this forms a dataset that can be further used in valuable ways.[94] A *labeled* instance includes *features* and a *label*.[95] Think of it as follows: {features, label} : (x, y).[96] The label is the thing we have already predicted or are trying to predict.[97] The features include one or more pieces of information about the world used as an input variable.[98] Data collection methods must be reliable from the beginning because bad data can easily be acquired through bad collection methods, which renders this data less reliable.[99] Reliability in this sense refers to how

---

90. IND. CODE OF JUDICIAL CONDUCT R. 2.13(a)(1) (noting that in "hiring court employees and making administrative appointments, a judge . . . shall exercise the power of appointment impartially and on the basis of merit").

91. *See* Natalie A. Pierce & Tiana R. Harding, *The Implications and Use of Artificial Intelligence in Recruitment and Hiring*, 62 ORANGE COUNTY LAW., 36, 38 (2020) (noting that "[h]istorical data used to train models also reflects past bias in HR decision-making . . . [, and] the potential for bias can still present concerns with regard to discriminatory impact, creating a legal risk").

92. Harini Suresh & John V. Guttag, *A Framework for Understanding Unintended Consequences of Machine Learning* (2020), v3, https://arxiv.org/pdf/1901.10002.pdf [https://perma.cc/94XS-JE42] (discussing and categorizing machine learning harms).

93. *Id.*

94. *Id.*

95. *Framing: Key ML Terminology*, GOOGLE MACHINE LEARNING CRASH COURSE, https://developers.google.com/machine-learning/crash-course/framing/ml-terminology [https://perma.cc/2XUK-NX7P] (last visited June 3, 2020).

96. *Id.*

97. *Id.*

98. *Id.*

99. *The Size and Quality of a Data Set*, GOOGLE DATA PREPARATION AND FEATURE ENGINEERING FOR MACHINE LEARNING https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality [https://perma.cc/9F9T-ZLFP] (last visited June 3, 2020).

trustworthy your data is.[100] Data may include labelling errors or noisy features, which are all issues of reliability with regard to data collection and use.[101]

Data Preparation (Preprocessing, Cleaning, and Labeling): Data from the world is usually not organized in ways that are easy to analyze (for example, changing non-numeric features into a numeric representation).[102] Thus, data preparation such as preprocessing, integrating, labeling, and other processes to clean the data may be applied to the data in the dataset.[103] In general, the preparation stage will most likely include splitting datasets into training data, testing data, and validation data.[104] Training data is used for model creation, testing data is used for model evaluation, and validation data is used for model adjustment.[105]

Model Development: Machine learning practitioners use the training data to (unsurprisingly) train the model.[106] In essence, machine learning practitioners provide all the answers for the model, which is built using the training data.[107]

Model Evaluation and Postprocessing: Machine learning practitioners use validation data to tweak the model, then use part of the data to test the model.[108] Machine learning practitioners know the answer (the answer being the label), but "test" the model to see if the model comes up with the same answer.[109] Of course, test data must not be previously used, it must be unseen data.[110] Once a model is evaluated and ready to be used, there may be post-processing steps (like simplifying or visualizing the data).[111]

Model Deployment: Model deployment is the act of transferring the model from the laboratory to real world use.[112] Once the model is in a real-world setting, there will likely be feedback that needs to be integrated into the model.[113] The input that the model sees after it is deployed may not look the same as the input it saw during training and evaluation, so further monitoring or training may be necessary to properly interpret that model's predictions.[114]

---

100. *Id.*
101. *Id.*
102. *Introduction to Transforming Data*, GOOGLE DATA PREPARATION AND FEATURE ENGINEERING FOR MACHINE LEARNING, https://developers.google.com/machine-learning/data-prep/transform/introduction [https://perma.cc/5MM8-ARD8] (last visited June 3, 2020).
103. Suresh & Guttag, *supra* note 92.
104. *Id.*
105. *Id.*
106. *Machine learning with structured data: Data analysis and prep (Part 1)*, GOOGLECLOUD SOLUTIONS, https://cloud.google.com/solutions/machine-learning/ml-on-structured-data-analysis-prep-1 [https://perma.cc/JZ54-ZHPK] (last visited Nov. 18, 2020).
107. Suresh & Guttag, *supra* note 92.
108. *Id.*
109. *Preparing your training data*, GOOGLECLOUD HOW-TO GUIDES, https://cloud.google.com/automl-tables/docs/prepare#split [https://perma.cc/8U3Q-L5YL] (last visited Nov. 18, 2020).
110. Suresh & Guttag, *supra* note 92.
111. *Id.*
112. *Id.*
113. *Id.*
114. *Id.*

V. PROBLEMS WITH ARTIFICIAL INTELLIGENCE

Bias, bias, bias.[115]

Artificial intelligence tools, generally speaking, are not absolutely perfect.[116] The following are examples of bias inherent in AI gleaned from Olteanu et al. (2019), Mehrabi et al. (2019), and Suresh & Guttag (2020). This is not an exhaustive list of all the biases that can creep into datasets and algorithms, but illustrates the major issues present in AI tools. Lawyers and judges should know these types of inherent imperfections in AI tools so they can understand the limits of these tools.

Aggregation Bias. Aggregation bias stems from the loss of detail when one (aggregate) model is (erroneously) used to describe a certain subgroup.[117] For example, experts know that diabetes patients in different ethnic groups have unique manifestations of complications.[118] Therefore, a single model to predict complications is unlikely to be appropriate for any one group.[119] Aggregation bias could compromise the doctor's appraisal of a patient, which may lead to medical errors.[120]

Algorithmic Bias. Algorithmic bias is a kind of umbrella term loosely used to describe situations where there is supposedly no bias in the input data, and the bias is created by the algorithm.[121]

Functional Bias. Functional bias refers to bias that is a result of platform-specific processes.[122] In the social media context, these "platform affordances" are driven by the platform's interests and are used to "nudge" users toward certain behaviors.[123] Using these behaviors creates distortions in datasets if unaccounted for.

Historical Bias. Historical bias relates to bias recorded in world history that ends up in data.[124] In this situation, models are trained on data that reflect

---

115. *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 68 (2019) (noting, for example, that "bias can surface in the context of input bias level (when the source data is biased because it may lack certain types of information), training bias (when bias appears in the categorization of the baseline data), or through programming bias (when bias occurs from a smart algorithm learning and modifying itself from interaction with human users or incorporating new data)").

116. *See generally* Brian L. Frye, *The Lion, the Bat & the Thermostat: Metaphors on Consciousness*, 5 SAVANNAH L. REV. 13, 18 (2018).

117. Suresh & Guttag, *supra* note 92.

118. *Id.*

119. *Id.*

120. *See generally* Gustavo Saposnik, Donald Redelmeier, Christian C. Ruff & Philippe N. Tobler, *Cognitive Biases Associated with Medical Decisions: A Systematic Review*, 16 BMC MED. INFORMATICS AND DECISION MAKING 138 (2016).

121. Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman & Aram Galstyan, *A Survey on Bias and Fairness in Machine Learning* (2019) at 6, https://arxiv.org/pdf/1908.09635.pdf [https://perma.cc/8NDP-RJCZ].

122. Alexandra Olteanu, Carlos Castillo, Fernando Diaz & Emre Kiciman, *Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries*, 2:13 FRONT. BIG DATA 1, 10 (2019).

123. *Id.*

124. Mehrabi et al., *supra* note 121, at 4 ("An example of this type of bias can be found in a 2018 image search result where searching for women CEOs ultimately resulted in fewer female CEO images due to the fact that only 5% of Fortune 500 CEOs were woman—which would cause the search results to be biased towards

historical inequities.[125] In one popular example, the use of word embeddings (one kind of natural language processing technique) trained on Google News articles were shown to exhibit gender stereotypes.[126] The model will predict reasonable analogies and is a useful tool involving natural language.[127] The model, however, will also answer "man is to computer programmer as woman is to *x*" with *x*=homemaker.[128]

Evaluation Bias. Recall that training data is used for model creation and testing data is used for model evaluation.[129] Evaluation bias occurs when the testing data does not adequately represent a certain population.[130] An example of this comes from common facial analysis testing datasets.[131] Just single digit percentages of the images in these testing datasets are of dark-skinned female faces.[132] The effect of this is that algorithms performing poorly in identifying this group are not penalized because the testing dataset failed to discover this poor performance.[133]

Measurement Bias. There are a number of kinds of measurement biases, which basically arise from how features are measured.[134] For example, the proxy variable "arrest" has been used to measure "crime" or "riskiness" in some recidivism risk prediction tools.[135] Because minority communities are highly policed and have higher arrest rates, these proxy variables are often mismeasured.[136] This has led to a different assessment of these groups, leading to higher false positive rates for those from minority communities.[137]

Misinformation and Disinformation Bias. Misinformation and disinformation bias refer to biases resulting from the intentional or unintentional spread of false information.[138] This false information can distort behavioral data.[139]

---

male CEOs. These search results were of course reflecting the reality, but whether or not the search algorithms should reflect this reality is an issue worth considering.") (internal citations and links omitted).

125. *Id.*

126. *See generally* Tolga Bolukbasi, Kai-Wei Chang, Venkatesh Saligrama & Adam Kalai, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*, https://arxiv.org/pdf/1607.06520.pdf [https://perma.cc/5JM5-MB4W].

127. *Id.*

128. *Id.* at 3. For an example of these analogies using word2vec with the Google News dataset, see Radim Rehurek, RARE TECHNOLOGIES, https://rare-technologies.com/word2vec-tutorial/#app [https://perma.cc/E2LV-8NFX] (last visited Nov. 18, 2020).

129. Suresh & Guttag, *supra* note 92.

130. *Id.*

131. *Id.*

132. *Id.*

133. *Id.*

134. Mehrabi et al., *supra* note 121, at 4.

135. Suresh & Guttag, *supra* note 92.

136. *Id.*

137. *Id.*

138. Olteanu et al., *supra* note 122, at 12.

139. *Id.*

Non-individual Agents Bias. This type of bias occurs when researchers misinterpret human behavior with organization or automated agent behavior.[140] Some interactions on social platforms are produced by organizations or automated agents.[141] Using this data to make inferences about human behavior will be distorted.

Normative Bias. Normative bias refers to biases "that are a result of written or unwritten norms and expectations of acceptable patterns of behavior on a given online platform or medium."[142] These norms can distort behavior.[143] Using or measuring these behaviors creates distortions in datasets if unaccounted for.

Omitted Variable Bias. Omitted variable bias occurs when variables that should be included in the model are excluded.[144] The effects of these missing variable(s) are then (erroneously) attributed to the included variables.[145]

Popularity Bias. "Popular" online items tend to be more visible.[146] However, an item's popularity can be engineered by, for example, fake reviews or social bots.[147] Ranking bias, one kind of popularity bias, occurs when top-ranked results affect our judgment simply by virtue of their top rank.[148] Social bias, another kind of popularity bias, occurs when content affects our judgment simply by virtue of the content of a social media feed.[149]

Population Bias. Population bias arises in the absence of proper randomization.[150] Characteristics of the population represented in the dataset are not representative of the target population that is intended to be analyzed.[151] Population biases affect the representativeness of a dataset.[152] For example, women may prefer to use social media platforms differently than men do.[153] Research has shown that certain groups are more represented on some social media networks compared to others.[154]

Presentation Bias. Presentation bias can lead to relevance and other kinds of judgments being made based largely on how the information is presented.[155]

---

140. *Id.*
141. *Id.*
142. *Id.* at 11.
143. *Id.*
144. Tomi Mester, *Statistical Bias Types explained (with examples) – part 1*, DATA36 (Nov. 17, 2020), https://data36.com/statistical-bias-types-explained/ [https://perma.cc/4UP3-9N7K].
145. *Id.*
146. Mehrabi et al., *supra* note 121, at 6.
147. *Id.*
148. *Id.*
149. *Id.*
150. *Id.* at 5.
151. *Id.*
152. Olteanu et al., *supra* note 122, at 6.
153. *Id.*
154. *See generally* Eszter Hargittai, *Whose Space? Differences Among Users and Non-Users of Social Network Sites*, 13 J. OF COMPUTER-MEDIATED COMM. 276, 276–97 (2007).
155. *See* Mehrabi et al., *supra* note 121, at 6.

Users can interact only with phenomena they actually perceive, so everything else is ignored.[156] Therefore, it could be the case that an individual does not see all the phenomena necessary to create detailed interactions.[157] We see this on the web when certain items are created, therefore only those items can be clicked.[158]

Redundancy. Redundancy refers to "[s]ingle data items that appear in the data in multiple copies, which can be identical (duplicates), or almost identical (near duplicates)."[159] Redundancy must be accounted for, otherwise it can distort data in datasets.[160] In the social media context, content redundancy often manifests through nonhuman accounts or multiple users or multiple entities posting collectively.[161]

Representation Bias. Representation bias can manifest when a population is underrepresented in a population and accounting for this underrepresentation when defining a sample from this underrepresented data source.[162] One study showed that image search results relating to occupation exhibit an under-representation of women.[163] For example, according to this study a search of an occupation with an equal number of women would have about 45% women in the search results.[164] Another study showed that facial analysis technologies had higher error rates for minorities (especially minority women), potentially due to dataset composition.[165]

Sampling Bias. Sampling bias arises when sampling is not random or focuses on one population.[166] Trends for one population are generalized to another population.[167] If not accounted for, this can distort data in datasets.

Simpson's Paradox. "Simpson's paradox can bias the analysis of heterogeneous data that is composed of subgroups or individuals with different behaviors."[168] In one example, after analyzing graduate school admissions data at UC Berkeley as a whole, it appeared there was a bias against women.[169] However, when admissions data was separated and analyzed based on individual

156. *See id.*
157. *Id.*
158. *Id.*
159. Olteanu et al., *supra* note 122, at 10.
160. *Id.*
161. *Id.*
162. Mehrabi et al., *supra* note 121, at 4.
163. *Id.*
164. Matthew Kay, Cynthia Matuszek & Sean A. Munson, *Unequal Representation and Gender Stereotypes in Image Search Results for Occupations*, CHI Conference Paper 3819, 3820 (2015).
165. Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, Conference on Fairness, Accountability, and Transparency (2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf [https://perma.cc/P73X-DTTF].
166. Mehrabi et al., *supra* note 121, at 5.
167. *Id.*
168. *Id.*
169. *See id.*

departments, women applicants were equal and in some cases had a small advantage over men.[170]

Temporal Measurement Bias. Temporal measurement bias arises when the time of an action is not properly taken into consideration creating distortions over time.[171] For example one study of comment length on a popular discussion website found that when looking at changes in comment length over time, researchers found that while the overall average in comment length decreased over time, users actually write longer comments as they survive.[172] Since late joiners write shorter comments, "their greater number leads to an instance of Simpson's paradox, where the overall average decreases while the series for each individual cohort increases."[173] This impacts patterns of observation and can distort data if not accounted for.[174]

Funding, Publication, Observer, and Cause-Effect Bias. When studies support the financial sponsors' interests, we call this funding bias.[175] When research results are purposefully exaggerated to get published, we can call this publication bias.[176] When researchers subconsciously project the researcher's expectations onto the research, we call this observer bias.[177] When researchers discover a correlation and conclude that it implies causation, we call this cause-effect bias.[178]

VI. DATASET DISCLOSURE FORMS AND MODEL DISCLOSURE FORMS

Advances in machine learning research and a deeper understanding of bias in AI systems make it clear that communicating the specifications of AI systems is imperative.[179] This is doubly important for lawyers and judges who need to have technical competence regarding AI system specifications to fulfill ethical rules. Lawyers and judges must require documentation of data provenance and individual machine learning models when using a specific AI tool.[180] In the absence of

---

170. *Id.*

171. *See* Olteanu et al., *supra* note 122, at 9.

172. Samuel Barbosa, Dan Cosley, Amit Sharma & Roberto M. Cesar-Jr, *Averaging Gone Wrong: Using Time-Aware Analyses to Better Understand Behavior*, 25th INT'L CONF. ON WORLD WIDE WEB (2016), https://arxiv.org/pdf/1603.07025.pdf [https://perma.cc/956M-43C7].

173. *Id.*

174. Olteanu et al., *supra* note 122, at 9.

175. Mester, *supra* note 144.

176. *Id.*

177. *Id.*

178. *Id.*

179. *See generally* Kate Crawford, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak, Varoon Mathur, Erin McElroy & Andrea Nill Sanchez, et al., *AI Now 2019 Report*, AI NOW INSTITUTE (2019), https://ainowinstitute.org/AI_Now_2019_Report.html [https://perma.cc/33VM-XS75].

180. Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 397 (2016) (recommending that legislation and related regulatory body be created so that "[c]ompanies seeking certification of an AI system would have to disclose all technical information regarding the product, including: (1) the complete source code; (2) a description of all hardware/software environments in which the AI has been tested; (3) how the AI performed in the testing environments; and

technology competency and mechanisms to disclose the inherent flaws in AI tools, these flaws can present grave consequences for relevant stakeholders.[181] In addition, AI tool developers should not just mention broad instances of flaws generally seen in AI, but disclose the specific facts that make these flaws material to the specific AI tool. Dataset Disclosure Forms and Model Disclosure Forms are examples of such documentation.[182]

There is no standardized process for documenting datasets and machine learning models used in AI tools, but there is no shortage of research attempting a model to explain AI systems.[183] This creates a problem for lawyers and judges who need to be able to understand and demonstrate the fairness of AI tools they are using. To address this gap, the author proposes the mandatory use of Dataset Disclosure Forms and Model Disclosure Forms. In most industries, products are often accompanied with information that describe "operating characteristics, test results, recommended uses, and other information" regarding the product.[184] Similarly, Dataset Disclosure Forms and Model Disclosure Forms should accompany datasets and models to inform dataset and model consumers.

Dataset Disclosure Forms and Model Disclosure Forms are documents that should accompany machine learning models and datasets. Model Disclosure Forms will provide details on model motivation, details, uses, performance, datasets, risks, distribution, and maintenance. Dataset Disclosure Forms will provide details on dataset motivation; composition; collection process; preprocessing, cleaning, and labeling; uses; distribution; and maintenance. Dataset Disclosure Forms and Model Disclosure Forms encourage transparent use of AI and related AI technology, increasing transparency in the legal process. Those lawyers and judges using AI tools must submit (or keep on file subject to audit) Dataset Disclosure Forms and Model Disclosure Forms in order to demonstrate compliance with applicable ethical rules.[185] To summarize, every AI tool used by an

(4) any other information pertinent to the safety of the AI. After disclosure, the Agency would conduct its own in-house testing to assess the safety of the AI program").

181. *See* Shlomit Yanisky-Ravid & Sean K. Hallisey, *"Equality and Privacy by Design": A New Model of Artificial Intelligence Data Transparency Via Auditing, Certification, and Safe Harbor Regimes*, 46 FORDHAM URB. L.J. 428, 453–55 (2019).

182. Crawford et al., *supra* note 179, at 8.

183. *AI Explainability Whitepaper*, GOOGLE, https://storage.googleapis.com/cloud-ai-whitepapers/AI%20Explainability%20Whitepaper.pdf (last visited Nov. 12, 2020) [https://perma.cc/6WJW-2WJZ] (noting that "[r]esearch progress in [explainable AI] has been rapidly advancing, from input attribution (LIME, Anchors, LOCO, SHAP, DeepLift, Integrated Gradients, XRAI etc.), concept testing/extraction (TCAV, DeepR, Towards Automatic Concept-based Explanations), example influence/matching (MMD Critic, Representer Point Selection, Influence Functions, Attention-Based Prototypical Learning), distillation (Distilling the Knowledge in a Neural Network, Distilling a Neural Network Into a Soft Decision Tree)") (internal links omitted).

184. Gebru et al., *supra* note 4, at 1.

185. This article makes use of Gebru et al., *supra* note 4, as a framework for the example Dataset Disclosure Form and Model Disclosure Form for lawyers and judges. The author would like to extend a special thanks, again, to the authors of this iconic article, including Timnit Gebru, who approved the use of a portion of their groundbreaking paper to develop the Dataset Disclosure Forms and Model Disclosure Forms.

attorney or judge, or in any way used in court proceedings, should be accompanied by Dataset Disclosure Forms and Model Disclosure Forms. An Example Dataset Disclosure Form and Model Disclosure Form will both be explored below.[186]

VII. DATASET DISCLOSURE FORMS

Dataset Disclosure Forms are necessary to ensure that all relevant parties are sufficiently educated regarding the contents of data in the dataset. Dataset Disclosure Forms also facilitate "greater reproducibility of machine learning results: without access to a dataset, researchers and practitioners can use the information in a [Dataset Disclosure Form] to reconstruct the dataset."[187] In the following, the author makes use of the example datasheet from the groundbreaking Gebru et al. (2020) datasheet and restyles and adapts it for use in the legal context. If the Dataset Disclosure Form is to be used in court, a representative must sign the Dataset Disclosure Form under penalties for perjury.

Motivation

Dataset creators must clearly articulate the reasons for creating the dataset and disclose all funding interests. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe the specific purposes of the dataset. "Purpose" should be broadly construed as the reason for which something exists, as well as an intended or desired result. Describe any specific tasks that were contemplated by the dataset. "Task" should be broadly interpreted as a piece of work expected of the data. Describe any specific gaps that needed to be filled. Answers should relate to the specific tool that is being developed, and not abstract goals like "to generate revenue" or "maximize shareholder profit."
- Name the entity (e.g., company, institution, organization) that created the dataset. Attach an organizational chart showing company ownership (including affiliated entities). List any affiliated entities with an interest (economic or otherwise) in any use or outcome of the use of the dataset (for example, if the use of the dataset relates to incarceration of individuals, disclose whether there is any entity in the ownership chain directly or indirectly engaged in the prison industry).
- Name and describe the internal team that created the dataset (e.g., team, research group). Describe whether the team has strategies in place to attract, develop, and advance a balance of race, sexual orientation, religion, age, gender, disability status, or any other dimension of diversity of the creators of the dataset. Attach an internal organizational chart (i.e., showing

186. Eran Kahana, *AI and the Law: 2019 in Review*, Stanford Law School (Dec. 31, 2019), https://law.stanford.edu/2019/12/31/ai-and-the-law-2019-in-review/ [https://perma.cc/5TD4-4M2K] (noting that standardizing explainable AI is an important effort because it promotes outcome-predictability and reliability).

187. Gebru et al., *supra* note 4, at 2.

responsibilities or reporting relationships of individuals or offices within the company).

- If an unrelated entity funded the creation of the dataset, name that entity. Attach an organizational chart showing affiliates and ultimate ownership of that entity. If there is an associated grant, provide the name of the grantor and the grant name and number.
- Describe whether development of the dataset was part of a research-based grant and whether the researchers have accepted restrictions on publishing this research for proprietary or national security reasons.
- List any other important details related to the motivations for creating the dataset.

Composition

These responses are intended to provide dataset consumers the information they need to make informed decisions about whether and to what extent they should rely on a dataset for a given task. You should not reveal information in violation of relevant privacy regulations in applicable jurisdictions. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe what the instances that comprise the dataset represent, making sure to indicate whether there are multiple types of instances.
- Indicate how many unique instances there are in total.
- Describe whether the dataset contain all possible instances or whether it is a sample from a larger set. If the dataset is a sample, describe the larger set and indicate whether the sample is representative of the larger set. If the sample is representative of a larger set, describe how this representativeness was validated or verified. If the sample is not representative of the larger set, describe why not. Describe whether and how representation or sampling bias has been accounted for.
- Describe the data in each instance. If there is a label or target associated with each instance, provide a description.
- If there is any information missing from individual instances, provide a description of this missing information and explain why this information is missing. This includes intentionally removed information. Describe whether any unintended effects have been accounted for by missing or removed information.
- If relationships between individual instances are made explicit, describe how these relationships are made explicit. If relationships between individual instances are not made explicit, describe why these relationships are not made explicit.
- If there are recommended data splits (e.g., training, development, validation, testing), provide a description of these splits and explain the rationale behind them. Describe whether any unintended effects have been accounted for by these splits.
- Describe any errors, sources of noise, or redundancies in the dataset including the anomaly detection techniques used. Describe whether any unintended effects have been accounted for by these anomalies.

- Indicate whether the dataset relies on external resources (e.g., websites, social media posts, other datasets). If so, a) indicate whether the external resources will exist, and remain constant, over time, or whether there are official archival versions of the complete dataset, and b) indicate whether there are any restrictions (e.g., licenses, fees) associated with any of the external resources that apply to a user. Provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate. Describe whether any unintended effects (like bias) have been accounted for by using those sources.
- Indicate whether the dataset contain data that might be considered confidential. "Confidential" data should be interpreted as broadly as possible. Examples of confidential data could be data that is protected by legal privilege or by doctor-patient confidentiality as well as data an individual simply intended to be kept secret. If so, provide a description.
- Indicate whether the dataset contain data that might be offensive, insulting, threatening, or might otherwise cause anxiety. If so, describe why.

The remaining questions in this section refer to datasets that relate to people.

- If the dataset identifies any subpopulations (e.g., age, gender), describe how these subpopulations are identified and provide a description of their respective distributions within the dataset. Describe whether any subpopulation should not be analyzed with the dataset because its representation in the dataset is not representative of the target population that is intended to be analyzed
- Indicate whether it is possible to identify or reidentify individuals through data linkage techniques, either directly or indirectly (i.e., in combination with other data) from the dataset. If so, describe how. If not, describe why not (e.g., explicit or quasi-explicit identifiers have been removed or generalized).
- Describe whether removing or generalizing identifiers have negatively impacted the dataset.
- Indicate whether the dataset contain data that might be considered sensitive (under the broadest interpretation) in any way (e.g., data that reveals names, gender, nationality, racial or ethnic origins, age, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history). If so, provide a description.
- List any other comments related to the composition of the dataset.

Collection Process

These responses are intended to provide dataset consumers the information they need to make informed decisions about whether and to what extent they should rely on a dataset for a given task. You should not reveal information in violation of relevant privacy regulations in applicable jurisdictions. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe whether the data associated with the dataset is new data, transformed legacy data, shared data, or purchased data (or a combination of these types). In any situation, describe how the data was validated or verified for quality and completeness.
- Describe in detail how the data associated with each instance was acquired. For example, describe whether the data was directly observable, reported by subjects, or indirectly inferred or derived from other data. If data was reported by subjects or indirectly inferred or derived from other data, describe how the data was validated or verified.
- Describe the mechanisms or procedures that were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API). Describe how these mechanisms or procedures are validated. Describe your internal compliance mechanisms that ensure the data was acquired lawfully. List the point of contact details for internal compliance issues.
- If the dataset is a sample from a larger set, describe the sampling strategy.
- Describe how representation bias, sampling bias, population bias, aggregation bias, and any other relevant bias issues are accounted for in the dataset.
- List the type of individuals involved in the data collection process (e.g., students, research assistants, contractors) and their means and amount of compensation. Describe whether the individuals involved in the data collection process have any interest (economic or otherwise) in any use or outcome of the use of the dataset and how this is confirmed.
- Explain the timeframe the data was collected compared to the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles). Describe how relevant time and place issues are accounted for in the dataset (e.g., regulatory or policy changes, court decisions, or market movements that may abruptly change data; postal boundaries are realigned).
- List all ethical review processes conducted (e.g., by an institutional review board). Provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

The remaining questions in this section refer to datasets that relate to people.

- Describe whether the data was extracted from the individuals directly, or obtained via third parties or other sources (e.g., via websites). If obtained via third parties or other sources, list the third parties or other sources and describe whether and to what extent the data was aggregated and what blending tools were used.
- Describe how the individuals were notified about the data collection. Provide a link or other access point to, or otherwise produce, the exact language of the notification.

- Describe how the individuals consented to data collection. Provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented. *Confirm that the consent was freely given, specific, informed or an unambiguous indication of the individual's wishes.*
- Describe whether the consenting individuals were provided a mechanism to revoke their consent in the future or for certain uses. Describe whether prior to giving consent, the individual was informed of this right to withdraw consent. Provide a description, as well as a link or other access point to the mechanism.
- Describe whether an analysis of the potential impact of the dataset and its use on data subjects (generally referred to as a data protection impact analysis or "DPIA") has been conducted. Summarize why you did or did not identify the need for a DPIA. If applicable, provide a description of this analysis, including the outcomes and any advice, as well as a link or other access point to any supporting documentation. List the point of contact details for the individuals who will review ongoing compliance with any DPIA or who will determine the need for a DPIA if one has not been conducted.
- List any other comments related to the collection of the dataset.

Preprocessing, Cleaning, and Labeling

These responses are intended to provide dataset consumers with the information they need to determine whether data has been processed in ways compatible with their chosen tasks. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe all preprocessing, cleaning, and labeling of the data. Describe any significant areas of concern in the unprocessed data like duplicates, missing data points, errors, inconsistencies, incompatible data formats, or data that would be considered messy or unbalanced.
- Provide a link or other access point to the raw, unprocessed data.
- List the point of contact details for the individuals who conducted the preprocessing, cleaning, and labeling of the data
- Provide a link or other access point to the software used to preprocess, clean, or label the instances.
- List any other comments related to the preprocessing, cleaning, and labeling of the dataset.

Uses

These responses are intended to provide dataset consumers with the information they need to determine whether data is or will be used in ways compatible with their chosen tasks. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe all known tasks the dataset has already been used for.
- Describe all tasks that the dataset could be used for. "Task" should be broadly construed as defined above including anything the user is interested

in seeing a response for (building a model, augmenting an already existing dataset, undertaking new analysis, etc.).
- List all tasks for which the dataset should not be used (such as decisions that impact people's lives and access to opportunities; high-stakes decision-making, law enforcement or the judicial process, etc.).
- Provide links or other access point to all papers, systems, or applications that use or have used the dataset.
- Describe any specific or cross-disciplinary knowledge necessary to make inferences from or use the dataset.
- Describe details about the composition of the dataset or the way it was collected and preprocessed, cleaned, or labeled that might impact the current use or any future uses. For example, you should describe anything that users might need to know to avoid uses that could result in unfair treatment of individuals or groups or other undesirable harms or biases. Include details about anything a user could do to mitigate these undesirable harms or biases (e.g., the use of fairness metrics or algorithms to mitigate bias in the dataset).
- Describe any copyright, intellectual property license, or terms of use of the dataset and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or terms of use, as well as any fees associated with these restrictions. Describe whether the whole dataset is subject to relevant licensing terms or terms of use or, for example, just one or more layers, data items, organizational structures, or metadata.
- Describe any limitation of liabilities, warranties, or other caveats or limitations relating to the use of the dataset and whether the dataset is provided "as-is."
- Describe any limitations on rights of the user to reproduce, modify, improve, use, transfer, or sell the dataset for any purpose.
- List all third-party imposed restrictions on the use of the data and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions. Describe whether the whole dataset is subject to relevant third-party imposed restrictions or, for example, just one or more layers, data items, organizational structures, or metadata.
- List any other comments related to the use or potential use of the dataset.

Distribution

These responses are intended to provide dataset consumers with information regarding distribution of the data. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe how the dataset is distributed.
- Provide the dataset's digital object identifier ("DOI") or similar persistent identifier. Describe the entity's association or membership with any organization in the research community such as DataCite.
- Describe all export controls or other regulatory restrictions that apply to the distribution of the dataset or to individual instances and provide a link or other access point to, or otherwise reproduce, any supporting

documentation. List the controlling agency and reasons for control, including any relevant Export Control Classification Number. Describe whether classification was through a formal commodity classification request or self-classified.
- List any other comments related to the distribution of the dataset.

Maintenance

These responses are intended to provide dataset consumers a plan for dataset maintenance. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- List all entities supporting, hosting, and maintaining the dataset and their means and amount of compensation. Describe whether the entities involved in supporting, hosting, and maintaining the dataset have any interest (economic or otherwise) in any use or outcome of the use of the dataset and how this is confirmed.
- List the point of contact details for dataset maintenance issues.
- Describe the process of monitoring and re-evaluating the dataset.
- Provide a link or other access point to erratum notices or other expressions of concern related to the data.
- Describe how often, by whom, and how updates will be communicated to dataset consumers.
- Describe any applicable limits on the retention of the data associated with the instances (e.g., individuals were told that their data would be retained for a fixed period of time and then deleted), including how these limits will be enforced.
- Describe how older versions of the dataset continue to be supported, hosted, or maintained. Describe how dataset obsolescence will be communicated to dataset consumers. Describe whether any obsolescence risk forecasting or lifecycle forecasting were conducted relating to the dataset or any of its versions.
- Describe how others can extend, augment, build on, or contribute to the dataset. Describe how these contributions will be validated or verified. Describe the process for communicating and distributing these contributions to other users.
- List any other comments related to the maintenance of the dataset.

Caveats and Recommendations

- List additional concerns that were not covered in the previous sections.

VIII. MODEL DISCLOSURE FORMS

Model Disclosure Forms are necessary to ensure that all relevant parties are sufficiently educated regarding the models and related algorithms. In the following, the author makes use of the example datasheet from the abovementioned groundbreaking Gebru et al. (2020) datasheet and restyles and adapts it for model use in the legal context. If the Model Disclosure Form is to be used in court, a representative must sign the Model Disclosure Form under penalties for perjury.

Motivation

Model creators must clearly articulate the reasons for creating the model and disclose all funding interests. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe the specific purposes of the model. "Purpose" should be broadly construed as the reason for which something exists, as well as an intended or desired result. Describe any specific tasks that were contemplated by the model. "Task" should be broadly interpreted as a piece of work expected of the model. Describe any specific gaps that needed to be filled. Answers should relate to the specific tool that is being developed, and not abstract goals like "to generate revenue" or "maximize shareholder profit."
- Name the entity (e.g., company, institution, organization) that created the model. Attach an organizational chart showing company ownership (including affiliated entities). List any affiliated entities with an interest (economic or otherwise) in any use or outcome of the use of the model (for example, if the use of the model relates to incarceration of individuals, disclose whether there is any entity in the ownership chain directly or indirectly engaged in the prison industry).
- Name and describe the internal team that created the model (e.g., team, research group). Describe whether the team has strategies in place to attract, develop, and advance a balance of race, sexual orientation, religion, age, gender, disability status or any other dimension of diversity of the creators of the model. Attach an internal organizational chart (i.e., showing responsibilities or reporting relationships of individuals or offices within the company).
- If an unrelated entity funded the creation of the model, name that entity. Attach an organizational chart showing affiliates and ultimate ownership of that entity. If there is an associated grant, provide the name of the grantor and the grant name and number.
- Describe whether development of the model was part of a research-based grant and whether the researchers have accepted restrictions on publishing this research for proprietary or national security reasons.
- List any other important details related to the motivations for creating the model.

Model Details

You should not reveal information in violation of relevant privacy regulations in applicable jurisdictions. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- State the current model version and include a detailed changelog that includes, but is not limited to, new features, known issues, security concerns, bias concerns, user interface improvements, bug fixes, retraining, etc.

- Describe the type of model, including its attributes, model architecture details, rules, and other detailed information about the type of model.
- List the contact details of the owner, curator, or manager of the model.

Uses

These responses are intended to provide model consumers with the information they need to determine whether the model is or will be used in ways compatible with their chosen tasks. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe the task that the model is used for in the present case.
- Describe all tasks that the model could be used for. "Task" should be broadly construed as defined above including anything the user is interested in seeing a response for (building another model, augmenting an already existing model, undertaking new analysis, etc.).
- Describe tasks that the model should not be used for (such as decisions that impact people's lives and access to opportunities; high-stakes decision-making, law enforcement or the judicial process, etc.).
- Provide links or other access point to all papers, systems, or applications that use or have used the model.
- Describe any specific or cross-disciplinary knowledge necessary to make inferences from or use the model.
- Describe details about how the model was developed that might impact current or future uses. For example, you should describe anything that the users will need to know to avoid uses that could result in unfair treatment of individuals or groups or other undesirable harms or biases. Include details about anything a user could do to mitigate these undesirable harms or biases.
- Describe any copyright, intellectual property license, or terms of use of the model and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or terms of use, as well as any fees associated with these restrictions.
- Describe any limitation of liabilities, warranties, or other caveats or limitations relating to the use of the model and whether the model is provided "as-is."
- Describe any limitations on rights of users to reproduce, modify, improve, use, transfer, or sell the model for any purpose.
- List all third-party imposed restrictions on the use of the model by users and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.
- List any other comments related to the use or potential use of the model.

Model Performance, Datasets, and Risks

- Summarize the model's performance and how performance may vary.
- Describe all datasets that were used to train, test, and evaluate the model. Attach a Dataset Disclosure Form for each dataset used to train, test, and evaluate the model.

- Describe whether the model will be able to learn from new data it encounters in the future.
- Describe any known risks or model use cases that may be harmful in any way. "Harmful" should be broadly construed as anything that can cause or could possibly cause harm to any person, entity, or process.

Distribution

These responses are intended to provide model consumers with information regarding distribution of the model. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- Describe how the model is distributed.
- Describe all export controls or other regulatory restrictions that apply to the model or to individual instances and provide a link or other access point to, or otherwise reproduce, any supporting documentation. List the controlling agency and reasons for control, including any relevant Export Control Classification Number. Describe whether classification was through a formal commodity classification request or self-classified.
- List any other comments related to the distribution of the model.

Maintenance

These responses are intended to provide model consumers a plan for model maintenance. The language used should be simple and easy to understand by a technologically competent layperson, not an expert.

- List all entities tasked with supporting and maintaining the model and their means and amount of compensation. Describe whether the entities involved in supporting and maintaining the model have any interest (economic or otherwise) in any use or outcome of the use of the model and how this is confirmed.
- List the point of contact details for model maintenance issues.
- Describe the process of monitoring and re-evaluating the model.
- Provide a link or other access point to erratum notices or other expressions of concern related to the model.
- Describe how often, by whom, and how updates will be communicated to model consumers.
- Describe how older versions of the model continue to be supported or maintained. Describe how model obsolescence will be communicated to model consumers. Describe whether any obsolescence risk forecasting or lifecycle forecasting were conducted.
- Describe how others can extend, augment, build on, branch, or contribute to the model. Describe whether the model can further develop through incremental learning or other learning to acquire new knowledge (with or without forgetting previously acquired knowledge). Describe how these contributions will be validated or verified and whether the model will be updated or retrained. Describe the process for communicating and distributing the contributions and retraining updates to other users.
- List any other comments related to the maintenance of the model.

Caveats and Recommendations

List additional concerns that were not covered in the previous sections.

CONCLUSION

This article describes the lawyer's duty of technology competence to advise their clients while discussing the ethical implications of using AI technologies in the lawyer's own legal practice. It also describes a judge's duty of technology competence to understand the legal and ethical challenges associated with AI, and the advantages and disadvantages of using or allowing the use of AI technology tools in their own courtroom. Fundamental rights are protected in no small part by lawyers and judges through the courts, which means competence with regard to AI technologies is of utmost importance for lawyers, judges, and other judicial officers. Lawyers and judges must understand what an AI tool can and cannot do while also understanding its effectiveness and biases when accomplishing its particular objectives. Lawyers and judges must therefore possess a higher level of competence when AI tools impact their practice. Education is at the center of this competency. An understanding of AI in the legal field starts with educating stakeholders about the fundamental aspects of AI, its challenges, and how to create frameworks for addressing these challenges. Dataset Disclosure Forms and Model Disclosure Forms are a necessary first step to ensure that lawyers and judges are sufficiently educated regarding the contents of data, models, and related algorithms. While Dataset Disclosure Forms and Model Disclosure Forms are only one transparency tool among many, they are necessary for lawyers and judges to uphold their ethical duties.