PROFESSIONAL RESPONSIBILITY CONSIDERATIONS IN AI FOR EDISCOVERY:

COMPETENCE, CONFIDENTIALITY, PRIVACY, AND OWNERSHIP

AI Ethics and Bias subteam of the Analytics and Machine Learning Project,
Published May 2023

TABLE OF CONTENTS

ABOUT THE AI ETHICS & BIAS PROJECT

Use of artificial intelligence ("AI") tools in eDiscovery creates new opportunities for attorneys. By extracting, analyzing, and applying information from large data sets, AI tools can provide new insights, systematize processes, speed time to resolution, and reduce costs. A notable example is technology-assisted review ("TAR"), a process that makes use of machine learning to prioritize or classify relevant material in document reviews. Legal practitioners may reduce costs, time, and mistakes by applying TAR in litigation, antitrust reviews, investigations, and other matters. However, as legal teams' uses of these technologies evolve, ethical issues may arise, particularly with the opportunities for reusing the results of the computer learning in future matters, but for different clients.

CONTRIBUTORS

The project team (organizations noted for identification purposes only) includes:

- Ricardo Baeza-Yates, Director of Research, Institute for Experiential AI at Northeastern University, USA (San Jose, CA)
- Lilith Bat-Leah, Vice President, Data Services at Digital Prism Advisors, Inc. (New York, NY)
- Darius Bennett, Darius Emeka Bennett, P.C., CEO and Attorney, Civil Litigation, eDiscovery and Criminal Defense (Birmingham, AL)
- Tara Emory, Senior Vice President of Strategic Growth and General Counsel at Redgrave Data (Falls Church, VA)
- David D. Lewis, Chief Scientific Officer at Redgrave Data (Denver, CO) [Trustee]
- Khrhysna McKinney, Principal at K L McKinney (Sugar Land, TX) [Trustee]
- Dana Bucy Miller, Associate Director, Legal Solutions, QuisLex Inc. (Baltimore, MD)
- James A. Sherer, Partner, BakerHostetler (New York, NY)
- George Socha, Senior Vice President of Brand Awareness, Reveal (St Paul, MN)

A. Introduction

Use of artificial intelligence ("AI") tools in eDiscovery creates new opportunities for attorneys. By extracting, analyzing, and applying information from large data sets, AI tools can provide new insights, systematize processes, speed time to resolution, and reduce costs. A notable example is technology-assisted review ("TAR"), a process that makes use of machine learning to prioritize or classify relevant material in document reviews. Legal practitioners may reduce costs, time, and mistakes by applying TAR in litigation, antitrust reviews, investigations, and other matters. However, as legal teams' uses of these technologies evolve, ethical issues may arise, particularly with the opportunities for reusing the results of the computer learning in future matters but for different clients.

TAR uses supervised machine learning,[1] where attorneys train the software by providing examples of documents that are or are not of interest, and the software builds a predictive model that finds more documents of interest. Some predictive models can be applied not just on the client matter from which training documents were chosen, but to new matters as well. Models can even be trained iteratively using attorney assessments of documents from a series of client matters, improving over months or years. Machine learning software is different from other legal technologies in two ways. First, the effectiveness of machine learning software can potentially improve as it is used. Second, that improved effectiveness is embodied in models that are separate from the software itself—models that potentially can be applied to new datasets, separate from those to which machine learning was applied.

A trained model is an unusual entity from a legal standpoint. It can be highly effective, but that effectiveness is often difficult to predict. It incorporates patterns learned from the judgment of reviewers, but has no true legal knowledge itself. It analyzes the text extracted from the data of a client (or clients), but is not client data itself. It is produced in an automated way by particular software, but is not software itself. It has its own economic value. These characteristics of AI models raise a set of novel questions discussed below, including competent use of the technology, implications for client confidentiality and privacy, and potential for claims of economic benefits.

[1] TAR workflows are often described as TAR 1.0 and TAR 2.0. TAR 1.0 workflows involve expert decisions over a finite set of documents to train the model that will be applied for decision making on the balance of the documents. TAR 2.0 or Continuous Active Learning ("CAL") workflows allow for a continuous improvement of the model based on each review decision for the duration of the review. TAR 1.0 produces a quality of decision-making based on the finite set of documents and associated decisions used for training the model. TAR 2.0 has the potential to improve results by updating the model with each decision made throughout the review.

B. Competence

As the most basic ethical duty for attorneys, the duty of competence means that attorneys who use AI must familiarize themselves with whether it is working as intended in a particular matter and then validate the results.[2] The American Bar Association's ("ABA") Model Rule of Professional Responsibility 1.1 sets forth the duty of attorneys to "provide competent representation to a client."[3] Rule 1.1's Comment 8, which requires attorneys to "keep abreast of changes in the law," was amended in 2012 to make explicit that such changes "includ[e] the benefits and risks associated with relevant technology."[4] While this was not a new duty, the ABA sought to remind lawyers that understanding the risks and benefits of technology can be essential to meeting the ethical duty of competence.[5] The ABA also noted that competence in technology is often key to protecting the confidentiality of client information.[6]

AI technologies pose challenges for this duty of competence. Considerations of the ABA guidance seem to center on two main issues: the knowledge, behavior, and competence of attorneys and the information they receive. For example, when using AI technology, attorneys need to appreciate that the effectiveness of the technology can vary greatly across matters and data sets. An attorney also needs to understand how information from clients and others may be incorporated into models created by machine learning, and the implications of that incorporation for the attorney's professional responsibilities, particularly around use of models beyond their stated purpose. These are discussed in the later sections of this document.

---

[2] The American Bar Association maintains a list of continuing legal education requirements for all 50 states, the District of Columbia, three US territories, and two Canadian provinces (https://www.americanbar.org/events-cle/mcle/) and 39 states have adopted requirements that attorneys maintain technological competence (https://www.lawsitesblog.com/tech-competence). *See also* MODEL RULES OF PRO. CONDUCT r. 1.1 ("A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation"); id. at cmt. 8 ("including the benefits and risks associated with relevant technology").

[3] Comment 2 to Model Rule 1.1 provides that "[a] lawyer can provide adequate representation in a wholly novel field through necessary study. Competent representation can also be provided through the association of a lawyer of established competence in the field in question." Model Rules of Pro. Conduct r. 1.1 cmt. 8.

[4] ABA Comm'n on Ethics 20/20, Resolution and Report to the House of Delegates for Resolution 105A (filed May 2012; adopted August 6, 2012), *available at* https://www.americanbar.org/content/dam/aba/administrative/ethics_2020/2012_hod_annual_meeting_105a_filed_may_2012.pdf. While not specific to machine learning, the concerns with emerging use of new technologies remain evergreen: "[s]ome forms of technology, however, present certain risks, particularly with regard to clients' confidential information. One of the objectives of the ABA Commission on Ethics 20/20 has been to develop guidance for lawyers regarding their ethical obligations to protect this information when using technology, and to update the Model Rules of Professional Conduct to reflect the realities of a digital age." *Id.* at 1.

[5] Id. at 3.

[6] Id. at 4, 12.

C. Confidentiality

AI tools can also raise ethical questions about whether and how confidential client information may be used. When AI models are used iteratively over a series of similar matters to develop and refine the model, this implicates ethical considerations, as some case specific information may be accessible to third parties through the model, and attorneys have a professional responsibility to refrain from disclosing information about their representation of clients.[7]

AI tools build models that refer to features (characteristics of data). Models in today's eDiscovery tools rely heavily on document features that include words, names, and phrases extracted from client communications and other files. Examining such a model may mean seeing important parts of the *content* of those files on which the model was trained, which could include sensitive client information. If a model is built off the data within a single matter, this usually does not pose an issue; however, if the model is ported from matter to matter, incorporating new data and feedback with each iteration, the model may then contain content that could be accessed by third parties examining the model.

For example, suppose a law firm regularly handles internal investigations and employment litigation related to harassment. An important task within such matters is finding (or ruling out) the presence of harassing emails within large data sets. An attorney who works on many such matters may develop great expertise in searching for these emails and be preferred for this work over one who has never handled such a matter before.

Similarly, a predictive model that was trained across data from many such matters may be more accurate than one built from scratch for a new matter. While using such a model may be highly cost-effective, it could, in some situations, later allow a third party who is able to examine that model to see client-specific vocabulary and personnel names that machine learning found to be predictors of harassing emails. The observer might be able to deduce which companies have been involved in harassment investigations, and even which employees may have been perpetrators or victims of harassment. Even if the model is not in a form that can be directly examined by a user, the same deduction may be possible by observing the model's behavior on a large set of documents.

A complex ecosystem of professionals—attorneys, legal service providers, consultants, software companies, and others—often is involved in the use of TAR and other AI technologies. Access to information preserved or collected for a legal matter is typically limited to attorneys working on that matter and those working under their direction. If an information artifact (such as a predictive model) contains confidential client information but is used in matters for other clients, attorneys should understand the risk of inappropriate access and their ethical obligations to protect against that. Attorneys who practice in accordance with bar and court admissions and who direct certain individuals to assist them still have ethical and professional responsibility for that work.

---

[7] *See* MODEL RULES OF PRO. CONDUCT r. 1.6 cmt. 2 ("This contributes to the trust that is the hallmark of the client-lawyer relationship"). As noted in the preamble to the Model Rules, "A lawyer should keep in confidence information relating to representation of a client except so far as disclosure is required or permitted by the Rules of Professional Conduct or other law. MODEL RULE OF PRO. CONDUCT pmbl., *available at* https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/model_rules_of_professional_conduct_preamble_scope/.

D. Privacy

Attorneys must also be cognizant of evolving privacy laws and regulations applicable to data consumed or produced by AI tools. As described in the confidentiality example above, models may contain personally identifiable information such as names, addresses, identifying account numbers, and more. Recent privacy regulations in several jurisdictions (including the GDPR for the European Union, CCPA/CPRA in California, CDPA in Virginia, CPA in Colorado, UCPA in Utah, and CTDPA in Connecticut) have expanded the rights of parties to inspect, delete, and control the use of their data held by an organization. Many organizations, including law firms and other legal service providers, have been grappling with their obligations under these regulations. What is less clear, however, is how these obligations apply if the original data is no longer held by the organization, but related data lives on in a predictive model.[8]

Consider an example where supervised learning has been used to train a model that predicts whether small business clients will win a lawsuit, and where the clients contractually require their data be deleted upon conclusion of litigation. This can pose a dilemma for further use of the model generated during the supervised learning. A firm might have a policy of preemptively deleting former small-business client data after some period to ensure privacy. What then happens if a client asks not only that their data be deleted, but that the data's effects on all predictive models be undone? This may technically be impossible unless data for all past clients is retained to allow the firm to retrain the model upon removing the requesting client's data. Such additional retention would lead a firm to increase, rather than decrease, the amount of client data held, and the attendant privacy risks (e.g., in a breach situation). If the data from customers or employees of the law firm's client is included, the issues expand to include data sourcing and consent considerations.

A variety of technical and process mitigation techniques may be employed, including automated omission of some types of textual features (e.g., proper names) from use by models, manual review and editing of models after training, cryptographic protections, and new forms of training algorithms. Regardless of these mitigations, however, questions remain about what laws and regulations may still be implicated, and what would be the associated responsibilities of legal practitioners.

[8] Anthony A. Ginart et al., *Making AI Forget You: Data Deletion in Machine Learning*, 33rd Conference on Neural Information Processing Systems (NeurIPS), 2019.

E. Ownership

Even if no private client information is included in a predictive model, an additional concern arises from the fact that the model may have economic value: who gets to capture and utilize that value? If a law firm or legal service provider derives value from a model trained on a client's data, does the client hold a stake? If a vendor further develops an internal model using client data, does the vendor hold a stake? Viewed from a different perspective, if the model is created within the confines of a particular matter at the direction of counsel, does its inclusion constitute attorney work product? If so, how does that factor into the reutilization analysis?

Returning to our harassment detector example from above, a law firm or alternative services legal provider (ALSP) may be able to conduct an investigation more quickly, and thus obtain more legal work, by using a model that was trained on data from past client matters. Further, the firm or ALSP might make the model available for licensing through a model marketplace (as several companies have set up), producing revenue directly associated with the model, not just with legal operations. Do or should past clients have claims on that revenue, or other intellectual property rights, in either scenario?

There are four types of data at issue in these scenarios: (1) client data; (2) training labels; (3) trained models; and (4) the predictions made by those models. In our harassment detector example, the client data are collected emails. Training labels are annotations made as to which of those emails relate to harassment and which do not. These might have been made by the law firm, an ALSP, a client, or any combination thereof. The trained model is the harassment detector, and the predictions are that model's output when applied to new email messages from other clients. The model, and its predictions, would not exist without both the client emails and the (possibly jointly created) training labels. Further, the results would not exist without the attorney work product.

Complicating this issue is the duty of attorneys to refrain from self-dealing in client representations. This includes the duty to avoid enriching oneself at the expense of the client and to avoid asking clients for gifts.[9] Where a lawyer benefits her practice by using client data to enrich her law firm, is advance client consent required?

Conversely, both clients and courts expect attorneys to learn from their legal work and apply that knowledge to future legal matters. Clients benefit from such efficiencies. Experienced attorneys command a premium on the market. Minimum amounts of experience are legally required in many contexts, and ongoing learning typically is an ethical obligation. Work done by experienced attorneys for one client—taking advantage of skills learned from working with other clients—is uncontroversial and expected. Work product, such as memoranda containing confidential information, is also understood to become a valuable commodity (though work

[9] *See* MODEL RULES OF PRO. CONDUCT r. 1.8(a) ("A lawyer shall not enter into a business transaction with a client or knowingly acquire an ownership, possessory, security or other pecuniary interest adverse to a client"); 1.8(b) ("A lawyer shall not use information relating to representation of a client to the disadvantage of the client unless the client gives informed consent"); 1.8(d) ("a lawyer shall not make or negotiate an agreement giving the lawyer literary or media rights to a portrayal or account based in substantial part on information relating to the representation"); 1.8(i) ("A lawyer shall not acquire a proprietary interest in the cause of action or subject matter of litigation the lawyer is conducting for a client"). Each of these sections has exceptions.

product may require sanitization of confidential information before use in other matters[10]). Attorneys are, however, expected to not use confidential client information in legal matters for other clients or for personal gain. The vast majority of attorneys have no difficulty maintaining the distinction between the two types of information related to past engagements, thus meeting both these ethical obligations.

The ethical boundaries are less clear with a predictive model trained by machine learning. A machine learning model does not "know" anything; it simply captures patterns in the data on which it has been trained. Unless special measures are taken, it will freely combine general characteristics of human language (e.g., that the presence of a particular profane word is predictive of an email message being harassing) with private information from particular clients (e.g., a code name of an internal project of one client where harassment was occurring). Indeed, there is not always a clear distinction between the two types of patterns. For example, a particular department at one client may have a greater predisposition toward harassment due to its work culture than a similar department at another client, and machine learning may latch onto client-indicating features. Therefore, special attention should be paid to maintaining confidentiality of client data when adapting existing models for new clients.

Thus, the data acquisition step is critical when considering model development approaches and strategy. Clients may be perfectly agreeable to the use of their information to benchmark and effectively "share" in service to battle-tested and more cost-efficient client deliverables, as long as it is done by consent. Whether or not that discussion occurs, attorneys and practitioners by extension should be cognizant of the original duty of care owed to the clients who provided the data, and they should consider the maxim of primum non nocere—first do no harm—even if use of the data and a model-derived approach could ultimately lead to a beneficial outcome. The practitioner should first do no harm vis-a-vis the client, its data, and the client's ultimate aims.

F. Conclusion

As discussed above, AI provides outsized opportunities to improve the efficiency and effectiveness of the practice of law. Nonetheless, AI also can produce potential landmines for the practitioner as new regulatory frameworks emerge and case law provides precedents on the appropriate application of AI and related technologies. Attorneys are well-advised to be aware of these shifting challenges as well as their attendant responsibilities.

[10]The Sedona Conference, *Commentary on Privacy and Information Security*, 17 SEDONA CONF. J. 1, 65 (2016) ("In instances where a client does not consent to retention of its confidential information after the close of a matter, the client file retained by the LSP may still contain work product that the LSP wishes to keep as precedent, form, or history (such as legal memoranda, pleading drafts, or case notes).[] Under these circumstances, the LSP should 'sanitize' those documents, removing confidential client information before storing the documents in the LSP's precedent bank or file repository.").