Professional Perspective

# AI, Pursuit of Justice & Questions Lawyers Should Ask

Julia Brickell, Columbia University, Jeanna Matthews, Clarkson University, Denia Psarrou, University of Athens & Shelley Podolny, Columbia University

Bloomberg Law

# AI, Pursuit of Justice & Questions Lawyers Should Ask

*Contributed by Julia Brickell, Columbia University, Jeanna Matthews, Clarkson University, Denia Psarrou, University of Athens & Shelley Podolny, Columbia University*

The acceleration of AI and automated decision-making systems in business, including the business of law, impels a close look at the potential impact of artificial intelligence on legal systems and the role of lawyers and judges within those systems. From facial recognition to gunshot tracking, from probabilistic genotyping to sentencing to discovery, the pervasive use of AI tools impacts society and the legal system in ways that are important to ascertain.

To protect the rule of law and the fundamental values it is intended to serve, it is necessary to understand the risks and benefits that AI and automated systems present, not in the abstract, but in the actual context of a particular use. For lawyers, this is a matter of both professional ethics and social morality, as misuse or misrepresentation (intentional or otherwise) can undermine both the perception and the reality of a legal system's functioning fairly, transparently, and without bias. Some harms—sentencing based on algorithms using biased data or a resume sent to the trash pile—cannot be remedied after the fact.

A number of industry and governmental entities have begun to articulate principles for the ethical use of AI systems. The American Bar Association issued Resolution 112, cautioning lawyers to recognize that competence is required to understand when the risk of AI outweighs its benefits. The U.S. government has launched initiatives to promote the trustworthy adoption and use of AI systems. The Council of Europe, through the European Commission for the Efficiency of Justice, has propounded an ethical charter on the use of AI in legal systems. The Institute for Electrical and Electronic Engineers (IEEE) is proposing multiple standards designed to build trust in AI systems.

## Model Rules Alone Cannot Protect the Justice System

Lawyers in the U.S. should increasingly expect to be held accountable for understanding the impact of the AI systems that they employ or any system on which they advise. This includes systems that they or their opponents use to generate discovery or evidence presented to a court. This also encompasses the need to address the accuracy and fairness of an AI system and its impacts on individuals and society.

The use of automated systems in sentencing is one striking example of the potentially dire consequences to the quest for a just legal system. As the 2016 case *Wisconsin v. Loomis*, 881 N.W.2d 749 (Wis. 2016) demonstrates, the use of automated decision-making systems in sentencing strikes at the heart of due process. In this case, the defendant was denied, on grounds of trade secret, an opportunity to understand an algorithm used by the prosecutor to rate his likelihood of recidivism. There was an absence of evidence of the weights and scores the tool assigned in making its recidivism evaluation but there was evidence that the algorithm was racially biased against people of color and in favor of whites. In addition, the system was based on a sample of national data not validated for applicability to Wisconsin, relied on group data not calibrated for assessment of an individual, and was developed to inform post-sentencing support, not for use in sentencing.

What would a defense attorney need to know to effectively convince a court not to admit for consideration in sentencing any result from a recidivism-assessment tool that is opaque, biased, and designed for a different purpose? What would a prosecutor need to be convinced of not to propose using the automated ranking in the first place? As AI systems increasingly appear in juror selection and sentencing or generate or "identify" evidence for civil and criminal trials and other legal processes, the societal importance of lawyers' competent use of AI systems is increasing.

## ABA Resolution 112 & Model Rules

ABA Resolution 112 specifically calls out the expectation that lawyers understand the risks of AI, including the risks of bias and harm to the legal system. Indeed, the ABA model rules of professional conduct can already be seen as holding lawyers accountable for understanding AI tools.

**Rule 1.1 Competence and Comment [8]** requires a lawyer to competently represent each client, "keep[ing] abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology." In the context of AI systems, competence should reasonably be understood to require knowledge and assurance, based on assessment by

 

a competent assessor, first, of the reliability of the AI system in producing fair, unbiased, and accurate results, and second, of the accuracy of the results actually obtained. It should also require an understanding of risks, including to security, privacy, and confidentiality (e.g. from access to or exfiltration of data), and potential risks and impacts of intentional or unintentional misuse on clients, opponents, customers, targets, bystanders, and the legal system as a whole.

To date, 39 states have addressed technical competence in connection with Rule 1.1. Similarly, codes of judicial conduct are beginning to address the risks of technology. See, Indiana Code of Judicial Conduct Rule 2.5, Comment 1 (competence requires knowledge and skill "including the benefits and risks associated with the technology relevant to the service as a judicial officer.")

**Rule 1.4 Communication** requires reasonable consultation with clients about methods and choices for accomplishing the client's objectives. That should include accurately communicating the availability, effectiveness, risk, and overall impact on costs of relevant AI systems, including obtaining competencies for effective operation. Depending on the intended use—e.g., e-discovery or regulatory reporting—the potential impact of inaccurate communication may be significant, and pose client risk.

**Rule 1.5 Fees** charges lawyers with responsibility not to demand unreasonable fees, necessitating consideration of AI's potential for accelerating work (such as by high recall, high precision document review). Conversely, unreliable outcomes (attributable, for example, to system flaws, lack of competent operators, or lack of fitness for the intended use), may incur manual re-work or even regulatory or court fines. And is it ever reasonable to charge for a lawyer's learning curve or inferior results from deploying technology in-house rather than engaging third-party expertise?

**Rule 1.6 Confidentiality** requires understanding the operation and security of AI systems to avoid unintended access to client information—e.g., by unsecure systems, "smart" assistants that transmit to the vendor, use of AI trained on a different client's data, open-source licenses that require sharing.

**Rules 1.7, 1.9 Conflict of Interest** may implicate reuse of AI systems trained on client data, or taking a trained technology to a new firm.

**Rule 2.1 Advisor** requires exercise of independent professional judgment, implicating lawyer understanding of the design, training, and operation of AI systems on whose outcomes the lawyer relies.

**Rules 3.3 (Candor Toward the Tribunal), 3.4 (Fairness to Opposing Party and Counsel) and 4.1 (Truthfulness in Statements to Others)** require sound information on the effectiveness (as operated) of AI systems that provide information on which a lawyer relies in making factual assertions—e.g., a prosecutor's sentencing memo containing algorithmic assessment of recidivism risk, a rule 26(g) representation of reasonable completion of production, assertions on reliability of evidence emanating from AI systems.

**Rules 5.1, 5.3 Supervision** impose on senior lawyers responsibility to supervise subordinates and third parties to ensure compliance with the rules of ethics; this in turn generates need for significant information on the trustworthiness of outputs from AI systems in order to supervise the uses of and representations about those outputs.

**Rule 8.4(d) Misconduct** requires avoidance of conduct prejudicial to the administration of justice. The responsible lawyer must not only know the effectiveness of AI systems whose output is offered to or used by the court, but also envision both the impact on participants in the justice system and the ability of the justice system to police for bias, prejudice, and other unintended consequences.

**Rule 8.4(g) Misconduct** prohibits professional conduct that a lawyer reasonably should know is discriminatory. Accordingly, a lawyer must understand potential biases of AI systems they employ—insight that the system designer or distributor may not have gathered. For example, hiring algorithms trained on the data of a non-diverse firm's current partners will likely select resumes reflecting similar schools, hobbies, zip codes, and accordingly candidates. An AI system that recommends prospective trial teams based on successes from prior non-diverse teams may well replicate their characteristics.

The appropriate implementation of these professional standards, while daunting in the age of AI, is fundamental to the protection of a just legal system. Indeed, in a justice system that depends on a full exchange of facts to derive a just result, effective and competent deployment of systems by which the facts are uncovered is fundamental to fairness and protection of the rule of law.

The impacts of an AI system may be difficult to discern. As a concrete example, we can look at the unforeseen and yet far-reaching impact of the Facebook algorithms' manipulation of users' social media feeds and the negative consequences for individuals and society. How can lawyers meet these ethical obligations? How can they assess and demonstrate the trustworthiness of an implemented AI system for a particular use? The expertise needed to understand AI systems—statistics, computer science, data science—is beyond the ken of the average lawyer.

## Role of Measurement & Transparency in Developing Trust

The solution lies in measurement and transparency. If a system is measured to be effective at meeting a task, operated by those with due competence, transparent, and fair in apportionment of accountability for explaining design, training, and implementation decisions—including to the designers, developers, and operators of the system—the ethical obligation can be met.

The frameworks mentioned at the top of this article align in this direction. While the specifics vary, a common focus is the promotion of the trustworthy adoption and use of AI systems. See e.g., United States government initiative Guidance for Regulation of Artificial Intelligence Applications (White House Office of Management and Budget, 2020) in which the first principle, "Public Trust in AI," expounds on need for trust in systems to be deployed, in view of risks, inter alia, to civil liberties, and the concomitant need for promotion of "reliable, robust, and trustworthy AI applications, which will support public trust in AI."

Trust in AI system use and operation is fundamental to support the rule of law, which requires trust in the recommendations and judgements of the legal system itself. Conversely, AI systems associated with opaque or biased decision-making undermine trust in the legal system; expediency is an inadequate justification for the damage they engender. Importantly, AI systems are often trained on historical data that reflects embedded historical biases that can promote further inequality. Without requirements for evaluations of trustworthiness, therefore, AI systems can institutionalize both intentional and unintentional biases, potentially augmenting abuse of power and undermining democratic ideals. See Global Governance of AI Summary Report 2018 Roundtable at 32.

A prime example is the use of the gunshot detection system ShotSpotter, which installs microphones to locate and identify the sound of gunshots. Police assert that they pick the neighborhoods to be surveilled based on where the most shootings take place, but the system is overwhelmingly employed in communities of color. Research has shown that ShotSpotter is regularly susceptible to false positives and that its placement only in particular neighborhoods inflates gunfire statistics. This in turn is used to justify heightened policing when, in fact, the validity and reliability of the technology has not been established. Similarly, AI predictive policing systems may perpetuate historical bias. Crimes are more likely to be detected where the systems are deployed, thus the practice, in both instances, reinforces historical bias under the guise of objective evidence.

## Evaluating the Trustworthiness of AI Systems

It is important to evaluate the use of an automated system in the operational context in which it is to be deployed. Developers of systems may offer generic testing results, but one should ask what evidence exists that the system will be effective and accurate in the desired use case. For example, probabilistic genotyping software, used to match evidence samples found at crime scenes to possible suspects, may be used in cases where the software has not been proven reliable, e.g., cases with a larger number of contributors or small evidence samples and cases involving multiple people who are genetically related. Similarly, e-discovery software may advertise 99% accuracy on some benchmarks, but the recall of relevant data in a particular case may vary substantially.

For valuable guidance as we seek to evaluate the trustworthiness of AI systems, lawyers may turn to the IEEE Ethically Aligned Design principles. The framework for "informed trust" is designed to be practically applicable to a variety of circumstances, and to enable lawyers to avoid the risks of untrustworthy systems being applied in the legal context in a manner that undermines trust in the immediate process or, more importantly, in the legal system as a whole. IEEE set out four basic principles to guide the ethical adoption of AI systems:

- **Effectiveness**. Solid information about the capabilities and limitations of an AI system to ensure fitness for the intended purpose.
- **Competence**. Certainty that operators have the skills and knowledge required for the effective operation of the AI system and adhere to those competency requirements.
- **Accountability**. Clear lines of responsibility to provide the rationale for decisions made in the design, development, procurement, deployment, operation, and validation of effectiveness for system outcomes.
- **Transparency**. Those affected by the output of an AI system have access to appropriate information about its design, development, procurement, deployment, operation, and validation of effectiveness.

To pursue these trust principles, lawyers may start their evaluation of an AI system by learning, for example:

- Under what conditions did the developers test the systems? How might the environment of the intended use differ?
- Is there research on bias and the potential for disparate impact in systems of this type? Are there key demographic groups for which the system was not tested? Might this system in this instance with this data have a disparate impact?
- How is effectiveness to be tested in the context of the current use? What samples need to be drawn?
- What are the competencies needed to evaluate the effectiveness of the AI system and to deploy it effectively?
- What evidence is there that the system has been tested on use cases similar to the proposed use?

Lawyers contemplating or confronting use of AI systems should seek evidence supporting these four trust conditions for information to evaluate the trustworthiness of the AI system in question. The greater the potential impact of the system's outputs—considering proposed use, potential impact of inaccuracies or bias on the rule of law or fundamental human values, likelihood of bias, pernicious impact of ostensible objectivity—the greater and more sound the evidence required to support its use.