

AI and Ethics Series

Bias and Fairness in AI Systems: Navigating Ethical Obligations in Legal Practice

As artificial intelligence tools become increasingly embedded in legal practice, law firms face critical questions about bias, fairness, and ethical responsibility. From trademark searches to case predictions, AI systems promise efficiency gains, but they also carry risks that attorneys must understand and actively manage. How can practitioners and law firms identify bias in AI systems, implement practical safeguards, and fulfill their ethical obligations to clients while harnessing AI's potential benefits?

Let's begin with a fundamental question: What types of bias should concern legal practitioners most? "The first step is understanding the level of bias risk," explains José Arenado, founder of Quo Vadimus Interactive. "If you point AI at a dataset of all registered trademarks just to index or retrieve text, the risk of bias is relatively low—it's essentially a database query. But when you ask AI to interpret or generate legal opinions, the risks rise significantly."

Melissa Ballesteros, Manager of AI and Legal Technology Adoption at Holland & Knight LLP, identifies specific categories of bias that attorneys should monitor, highlighting three primary concerns:

1. Training data bias. If the AI is trained on historical outcomes, it can reinforce old patterns, including biases in trademark approvals or rejections.
2. Language and cultural bias. AI can miss nuance in translation or context, which is a real issue in cross-border trademark matters.
3. Algorithmic fairness. Some systems favor certain industries or client profiles, and that's not always easy to spot unless you're looking for it.

Rule 8.4(g) of the American Bar Association's Model Rules of Professional Conduct prohibits lawyers from engaging in discriminatory conduct, raising the question of whether biased AI systems could expose firms to liability. Both experts confirm this risk. Arenado emphasizes that "the safeguard is simple: verify," while acknowledging that "yes—if unchecked," biased AI can create liability.

Ballesteros outlines the potential consequences including "financial liability" and "loss of competitive edge" more extensively: "Yes, biased AI can expose firms to serious risks: Discrimination claims if outputs disadvantage certain groups. Reputational damage if clients lose trust."

These risks underscore that AI adoption is not merely a technology question, but a risk management imperative that implicates the firm's ethical standing, client relationships, and market position. In fact, critical competence lies not merely in identifying bias after the fact, but in knowing when AI is appropriate at all.

Since Model Rule of Professional Conduct 1.1 requires lawyers to provide competent representation, and Comment 8 explicitly specifies staying informed about relevant technology [1], Ballesteros advocates for comprehensive organizational approaches since this requirement applies to individual attorneys as well

as firm-wide systems and culture. "If you're using AI in your legal work, you need to understand it. That means knowing what the tool does and where it might misfire."

Arenado draws an analogy to a familiar experience: "Anyone who has run a search knows how a small typo can return completely irrelevant results. That's not bias—that's noise. Attorneys should be able to distinguish low-risk uses...from high-risk ones."

The duty to provide diligent and zealous representation under Model Rules 1.3 and 1.7 raises questions about when biased AI tools might compromise client service. Arenado frames this as a matter of proper boundaries: "Using AI to gather case law or review trademark filings is straightforward and can even enhance efficiency. But using AI to generate strategic advice risks embedding hidden bias into your client's legal strategy. The principle is simple: keep AI in the proper lane, and client interests remain protected."

Ballesteros illustrates how bias translates into tangible representation failures: "Biased AI doesn't just create inefficiencies. It can compromise representation. For example, [by] delivering weaker research for certain case types, recommending strategies based on client traits instead of legal merit, misvaluing IP assets based on flawed historical data." She identifies the ethical tipping point: "This becomes an ethical issue when attorneys rely on AI without verifying its output. They miss signs that the AI is off base for a particular client. Service quality varies across demographics and no one catches it."

The attorney's professional judgment must remain paramount. As Ballesteros concludes: "We owe our clients diligence and zeal. That means knowing when to challenge the machine."

Both Ballesteros and Arenado advocate for transparency. Arenado states: "Disclosure should be the norm. Many attorneys don't realize that junior associates and staff are already using AI tools in their daily workflow. Clients deserve transparency when any tool meaningfully shapes their work product—just as they would if outside research or third-party services were involved."

Ballesteros agrees and emphasizes that disclosure should be meaningful rather than overwhelming: "Disclosure doesn't mean a tech lecture. It means explaining how AI supports the users so they can produce more valuable work. Flagging any limitations that matter. Clarifying how client data is handled when AI is involved."

This consensus on transparency reflects a broader principle: Clients have a right to understand the tools shaping their legal representation, particularly when those tools carry inherent limitations or risks.

Moving from principle to practice, Arenado advocates for simplicity and verification: "Bias mitigation begins by keeping AI focused on what it does well. Don't ask it to do two things at once. Break tasks into single, verifiable steps. Repeat the process until the output is accurate. The more complexity you pile on, the more likely the system is to introduce errors or drift into its own assumptions." He adds that attorneys should "treat AI like any other research source. Cross-check, validate, and take ownership of the final product. Responsibility never shifts to the tool."

Ballesteros offers a more structured governance framework based on her experience implementing AI systems at a global firm: "AI governance protocols. Appoint ethics leads, require approval for new tools, and standardize risk evaluation criteria. Bias audits. Use diverse datasets, compare outcomes across

groups, and document findings. Multi-level review. Define when partner review is required, use checklists to validate AI work, and sample outputs for quality control."

Specifically, she recommends the following institutional measures: "Mandatory AI literacy training for all attorneys. Clear documentation on use cases and limitations for each tool. Community building between users with different technical expertise can help each other create better work product that addresses deeper questions. A living inventory of AI tools and their profiles."

She also emphasizes the importance of diverse perspectives in testing: "Include evaluators from different backgrounds and practice areas. Reflect client industry diversity." This approach recognizes that bias often manifests differently across contexts and that homogeneous testing teams may miss critical issues. Additionally, she recommends maintaining verification documentation. Record how outputs were validated. Note when and why attorneys override AI recommendations."

Effective bias mitigation requires ongoing attention rather than one-time audits. The tools evolve, use cases expand, and new risks emerge, making continuous monitoring essential.

Looking ahead, both experts envision increased regulation. Arenado advocates for use-case-focused regulation: "Regulation should focus less on the technology itself and more on the use-cases. Guardrails are needed for high-risk applications: Prohibit AI from issuing legal opinions, require human review for risk assessments, and mandate disclosure when tools shape client advice." His approach centers on preserving human judgment: "Attorneys must remain the final authority."

Ballesteros anticipates more comprehensive regulatory developments: "Certification standards for legal AI tools, likely from bar associations. Mandatory bias testing before deployment. Stronger disclosure rules around AI use. Competency requirements for attorneys using advanced AI." She foresees that "future ethics rules will likely include: Clear AI oversight responsibilities. Required bias audits and documentation. Transparency standards for clients and courts. Defined accountability between lawyers and tech vendors."

Both experts urge proactive rather than reactive approaches. As Ballesteros advises: "Don't wait for regulators to tell you what to do. Build governance now. Lead the way, trying to be the best human you can be."

Both practitioners share an optimistic yet grounded vision of AI's role in legal practice. Arenado offers perspective amid the hype: "The real benefit of AI is not to replace an attorney's work, but to allow attorneys to do their very best work with less stress and greater clarity." His conclusion emphasizes the enduring centrality of human judgment: "AI won't replace the attorney's judgment anytime soon—and very likely not within the span of your career. What it can do, when used wisely, is strip away the noise so you can focus on what matters most: exercising your best judgment. That is the true promise of AI in law: not stress, but clarity."

Ballesteros's approach centers on the human dimension of technological change: "I help turn AI adoption into a human experience where people feel empowered, connected, and ready to grow." This philosophy recognizes that successful AI integration requires not just technical competence but cultural transformation—building trust, fostering collaboration, and maintaining the human connections that define excellent legal service.

The path forward requires attorneys to embrace AI's capabilities while maintaining critical distance, to leverage efficiency gains while preserving professional judgment, and to adopt new tools while upholding timeless ethical obligations. Bias and fairness in AI systems are not problems to be solved once and forgotten, but ongoing responsibilities that demand vigilance, transparency, and commitment to equitable outcomes. As law firms navigate this evolving landscape, the practitioners featured in this article remind us that technology should serve justice, not supplant it, and that the attorney's duty to clients, competence, and fairness remains unchanged, regardless of the tools employed.

****Contributor Bios**** These quotes reflect the contributors' personal perspectives and are not tied to their official positions.

José Arenado is the founder of Quo Vadimus Interactive (QVI), a boutique digital studio that builds AI-powered legal tech solutions for trademark and IP law firms worldwide. With over twenty years of experience in software engineering and digital strategy, José helps firms adopt technology responsibly—enhancing efficiency while preserving the judgment and integrity that define great lawyering. Learn more at quovadimusinteractive.com.

Melissa Ballesteros serves as Manager of AI and Legal Technology Adoption at Holland & Knight LLP, where she guides global teams through technological transformation with clarity and empathy. With twenty years of experience across legal technology and international industries, she brings people and technology together through trust, teamwork, and purposeful change. Melissa is co-founder of Rockstar Women in AI, a global community elevating women into the future of AI, and founder of Creative Unwind, a collaborative art experience fostering team connection.

[1] Model Rules of Prof'l Conduct r. 1.1 cmt. 8 (Am. Bar Ass'n 2012).

Author:

Lauren Adler, Alt Legal (United States)