



The Human Behind the Machine: Bias and Accountability in Legal AI

California Minority Counsel Program - 2026 Annual Business Conference

Date & Time:	September 25, 2026 from 2:45 PM to 4:00 PM
Title:	The Human Behind the Machine: Bias and Accountability in Legal AI
Format:	75-minute moderated panel program, informed by an opening audience poll
CA MCLE:	1.25 hours in Implicit Bias / Bias-Reducing Strategies
Panel:	Moderator: Amy Chang (Executive, CAA) Panelists: David Wang (CTO, Cooley), Ally Bennet (Deputy GC, OpenAI), and Majorie Soto-Garcia (Partner, MoFo)

Description

As law firms and legal departments increasingly adopt AI tools, lawyers must understand how implicit, explicit, and systemic bias can be embedded in AI-assisted legal workflows. This program examines how bias can enter AI systems through training data, algorithm design, deployment decisions, feedback loops, and human overreliance on automated outputs. Panelists will discuss how AI bias can affect protected groups in contexts such as legal research, document review, compliance, investigations, hiring, legal operations, risk assessment, and employment decision-making. The panel will also provide practical bias-reducing strategies lawyers can use when evaluating AI tools, designing governance structures, supervising AI-assisted work, and managing vendors.

Learning Objectives

- Identify how implicit human bias and systemic bias can enter AI-assisted legal workflows through data, design, deployment, feedback loops, automation bias, and confirmation bias.
- Recognize and assess how AI-assisted legal, employment, compliance, or investigative decisions may create or amplify disproportionate harm to protected groups.
- Evaluate a legal AI use case using a bias-aware framework that addresses affected stakeholders, disparate effects, transparency, testing, meaningful human oversight, escalation, and accountability.
- Apply bias-reducing strategies, including use-case limits, objective prompts and criteria, documented human review, cross-functional governance, vendor diligence, monitoring, remediation, and suspension of inadequately supervised tools.

Moderator and Speakers

- Moderator: Amy Chang (Executive, CAA)
- Speaker: Ally Bennet (Deputy GC, OpenAI)
- Speaker: David Wang (CTO, Cooley)
- Speaker: Majorie Soto-Garcia (Partner, MoFo)

Panel Roadmap and Discussion Prompts

Subtopic 1 - Opening, Framing, and Audience Poll

The moderator will frame AI as a socio-technical system that reflects choices made by people, organizations, vendors, and institutions. The opening poll will help tailor the discussion to the audience's actual experience with AI adoption.

- *Has your organization adopted AI tools for legal work?*
- *Are you personally using AI for legal research, document review, drafting, investigations, compliance, or legal operations?*
- *How confident are you that your organization has evaluated whether its AI tools could produce disparate outcomes for protected groups?*

Subtopic 2 - How Bias Enters Legal AI

Panelists will discuss how bias may arise across the AI lifecycle and how those risks can affect legal and employment-related decision-making.

- **Training data:** Historical datasets may reflect existing societal, workplace, and legal-system inequities.
- **Algorithm design:** Optimization choices may prioritize efficiency, cost, or pattern recognition without adequate disparate-impact testing.
- **Deployment decisions:** A tool may create different risks depending on whether it is used for low-risk drafting assistance or high-impact decisions involving people's rights, employment, or access to services.
- **Feedback loops:** Biased outputs can generate downstream data that reinforces the original bias over time.
- **Human implicit bias:** Confirmation bias and automation bias can cause lawyers to over-credit AI outputs or fail to scrutinize unsupported recommendations.
- *Where does bias enter the AI pipeline, and who is most likely to be harmed?*
- *How does implicit bias in human decision-making compound AI bias when lawyers review or rely on AI-generated outputs?*
- *What examples show AI producing or amplifying discriminatory outcomes in legal, employment, or compliance contexts?*

Subtopic 3 - Evaluating AI Through a Bias-Aware Responsible AI Lens

Panelists will provide a practical framework for evaluating whether a particular AI tool is appropriate for a particular legal use case.

- **Purpose and fit:** Is the tool appropriate for this task, and could errors disproportionately harm protected groups?
- **Use-case boundaries:** What uses are approved, restricted, or prohibited?
- **Prompt design and criteria:** Are users asking for objective assistance, or are they inviting subjective judgments about credibility, intent, professionalism, risk, or performance?
- **Affected stakeholders:** Who is affected, and were their interests considered before deployment?
- **Transparency and explainability:** Can the legal team explain the output well enough to supervise, challenge, or defend reliance on it?

- Testing and escalation: Is there a process for disparate-impact testing, human review, and escalation when outputs appear biased or unexplained?
- Oversight: Who is accountable for ongoing monitoring and remediation?
- *Before adopting or relying on an AI tool, what questions should legal teams ask to surface bias risks for protected groups?*
- *When should a lawyer narrow an AI use case or decide that AI should not be used for a particular decision?*
- *What is the lawyer's professional obligation when the lawyer suspects, or reasonably should suspect, that an AI tool is producing biased outcomes?*

Subtopic 4 - Governance Strategies to Prevent Discriminatory Outcomes

This portion of the panel will focus on accountability structures that help prevent AI tools from amplifying implicit bias or creating discriminatory outcomes.

- Cross-functional AI review committees involving Legal, Compliance, IT, HR, DEI, privacy, security, and business stakeholders where appropriate.
- Bias or equity impact assessments before high-impact deployment.
- Documentation of approved uses, prohibited uses, known limitations, and escalation protocols.
- Defined human-review points for outputs affecting employment, legal rights, access to services, or other significant interests.
- Training users to recognize automation bias, confirmation bias, and unsupported AI-generated conclusions.
- Incident response plans for biased or discriminatory outputs, including suspension, remediation, and documentation.
- *Who should own AI bias prevention in a legal department or law firm?*
- *What governance structures help ensure AI supports rather than replaces professional judgment?*
- *How can organizations create a culture where attorneys feel empowered and obligated to challenge AI recommendations?*

Subtopic 5 - Vendor Management and Contracting Guardrails

Panelists will discuss vendor diligence as a bias-prevention strategy, while recognizing that some vendors may not provide complete transparency into training data, model design, or independent bias testing.

- Ask what the vendor can disclose about training data, validation data, model limitations, and intended uses.
- Ask whether the tool has been tested for disparate impact, by whom, and whether the testing matches the client's intended use case.
- Seek documentation sufficient to evaluate outputs, limitations, and risk controls.
- Negotiate data-use restrictions, including limits on using client data for model training or improvement without informed consent.
- Seek audit or inspection rights where feasible, particularly for high-impact use cases.
- Preserve remediation, suspension, and termination rights when the tool produces biased outcomes or the vendor fails to meet agreed safeguards.
- Monitor post-deployment outputs and re-evaluate the tool when the model, use case, data, or legal requirements change.
- *What diligence questions should legal teams ask every AI vendor about bias and equity?*
- *What contracting guardrails can reduce the risk of discriminatory outcomes?*
- *If the vendor cannot provide meaningful transparency, testing, or audit capability, when should the legal team recommend not using AI for the proposed use case?*

Subtopic 6 - Practical Takeaways

- *What is one practical step lawyers should take tomorrow to reduce the risk that AI tools amplify implicit bias or produce discriminatory outcomes?*
- Evaluate the use case before evaluating the tool.
- Do not use AI as the sole decision-maker in high-impact contexts affecting people's rights, employment, or access to legal or other significant services.
- Require human review that is trained, meaningful, and documented.
- Ask whether the tool could disproportionately affect protected groups.
- Watch for confirmation bias and automation bias in your own review of AI outputs.
- Document the purpose, limitations, approved uses, prohibited uses, and escalation path for AI tools.
- If you cannot explain, test, audit, or meaningfully supervise the tool, reconsider whether it should be used.

California Implicit Bias MCLE Qualification Summary

This program is designed to qualify for 1.25 hours of California MCLE credit in Implicit Bias / Bias-Reducing Strategies by addressing the recognition and elimination of bias, the impact of implicit and systemic bias on protected groups, and practical steps lawyers can take to recognize and address their own implicit biases when using or advising on AI tools.

CA Requirement	How This Program Addresses It
Identifying and understanding bias	Subtopic 2 identifies how bias enters AI systems through training data, algorithm design, deployment decisions, and feedback loops. It also explains how implicit bias in human decision-making can compound AI bias through confirmation bias, automation bias, and uncritical reliance on unexplained outputs.
Impact on protected groups	The program addresses how AI bias can disproportionately affect communities defined by race, ethnicity, gender identity, sexual orientation, national origin, disability, age, socioeconomic status, and other protected characteristics. The opening audience poll allows the moderator and panelists to tailor examples to the audience's experience with legal research, drafting, document review, investigations, compliance, hiring, legal operations, and other equity-sensitive uses.
Strategies to recognize, reduce, or eliminate bias	Subtopics 3 through 6 provide practical strategies for reducing bias in AI-assisted legal work, including responsible AI review, use-case boundaries, human oversight, prompt discipline, escalation pathways, cross-functional governance, vendor diligence, contracting guardrails, audit rights where feasible, post-deployment monitoring, and concrete takeaways lawyers can apply immediately.

Citations and Resources

California MCLE / Implicit Bias Authority

- California MCLE Requirements, State Bar of California.
- Cal. Bus. & Prof. Code § 6070.5.
- Rules of the State Bar, Title 3, Division 5, Chapter 1, Rule 3.600 et seq.

California Rules of Professional Conduct / California Attorney Duties

- California Rule of Professional Conduct 1.1 — Competence.
- California Rule of Professional Conduct 1.6 — Confidential Information of a Client.
- California Rule of Professional Conduct 3.3 — Candor Toward the Tribunal.
- California Rules of Professional Conduct 5.1 and 5.3 — Responsibilities of Managerial/Supervisory Lawyers and Responsibilities Regarding Nonlawyer Assistance.
- Cal. Bus. & Prof. Code § 6068(d), (e).

ABA Model Rules / ABA Guidance

- ABA Model Rule 1.1, Comment 8 — Technology Competence.
- ABA Formal Opinion 512, Generative Artificial Intelligence Tools (July 29, 2024).
- ABA Formal Opinion 477R, Securing Communication of Protected Client Information (May 11, 2017; revised May 22, 2017).

California Bar AI Guidance

- State Bar of California, Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law.
- State Bar of California, Artificial Intelligence Ethics & Technology Resources Page.

Cases / Sanctions Involving AI Hallucinations

- *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023).
- *Lacey v. State Farm Gen. Ins. Co.*, No. 2:24-cv-05205-FMO-MAA, 2025 WL 1363069 (C.D. Cal. May 5, 2025).

Technology / Governance Resources

- NIST Artificial Intelligence Risk Management Framework 1.0.
- NIST AI 600-1, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (July 2024).
- EEOC Artificial Intelligence and Algorithmic Fairness Initiative.
- DOJ/ADA.gov, Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring.
- California Civil Rights Department / Civil Rights Council Regulations on Employment Discrimination Related to AI and Automated-Decision Systems.
- NYC Local Law 144, Automated Employment Decision Tools.
- Colorado SB24-205, Consumer Protections for Artificial Intelligence.

Timing Summary

Segment	Topic	Time
1	Opening, framing, and audience poll	5 minutes
2	How bias enters legal AI	15 minutes
3	Bias-aware responsible AI review	15 minutes
4	Governance strategies	15 minutes
5	Vendor management and contracting guardrails	15 minutes

6	Practical takeaways and closing	10 minutes
Total		75 minutes / 1.25 hours