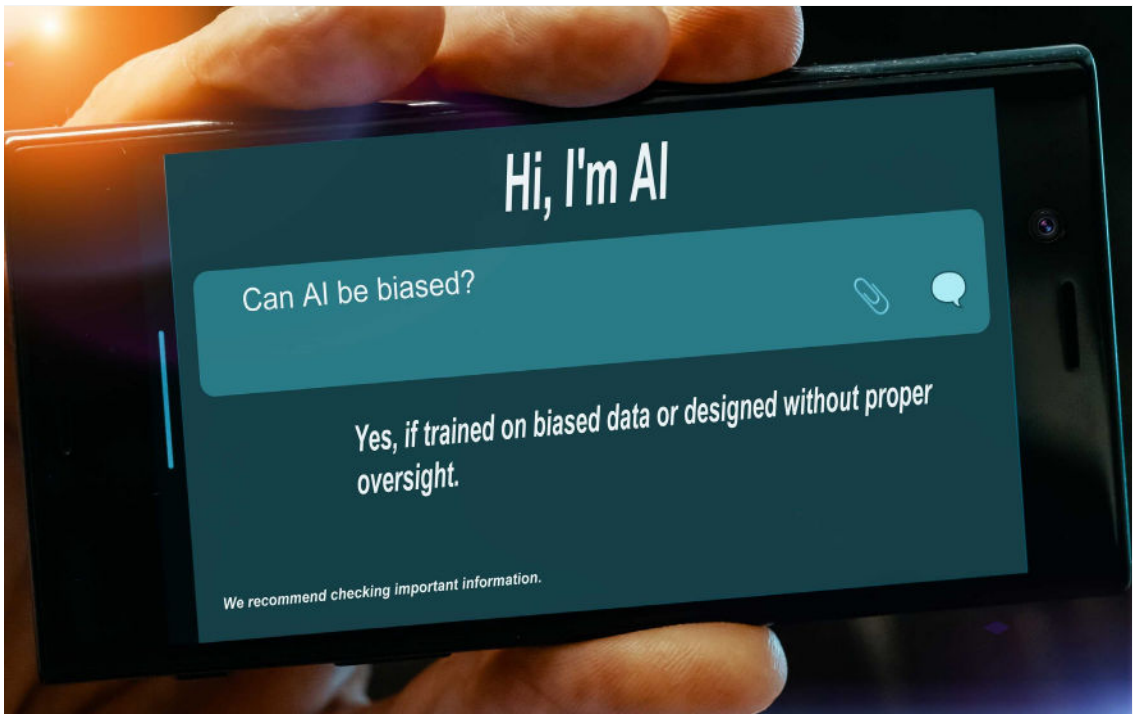
LOS ANGELES & SAN FRANCISCO
Daily Journal
www.dailyjournal.com
WEDNESDAY, AUGUST 12, 2026

PERSPECTIVE | MCLE

Bias is implicit in all AI, even legal AI

Attorneys who use AI must do so with their eyes wide open, understanding its limitations and recognizing that their work is just beginning whenever AI has been used.

By Susan Greenberg



Shutterstock

Artificial Intelligence (AI) is the topic of the day. From copyright infringement claims to hallucinations, it's impossible to bypass the subject. The legal profession, along with medicine, engineering, the arts and just about every other area of human endeavor, is finding new and interesting ways to utilize nonhuman intelligence.

But AI was created by humans and, despite best efforts, it can't help but reflect human biases and flawed perspectives. Human biases are encoded into machine learning algorithms, adversely and inadvertently impacting protected classes. Bias can find its way into algorithmic decisions via proxies for race, sex or age; they can show up through choices of definitions and features.

Experts in just about every discipline have been working overtime to keep up with the practical and ethical implications of the technology, and laws have been enacted to mitigate the potential for inequitable treatment of individuals and groups. Court cases have examined the bias inherent in tools such as language learning models (LLMs) and generative AI.

Much attention has focused on implicit bias in AI tools used by corporate HR departments for tasks such as recruiting and hiring. Evidence shows that without intending to, businesses have closed the door on promising candidates or denied them advancement because of implicit bias within the AI tools. Local Law 144, adopted by New York City in 2021, requires companies that use automated employment decision tools (AEDT) to screen candidates to conduct an annual audit to evaluate the AEDT's potential disparate impact based on protected demographic categories.

Far less attention has been paid to implicit bias within AI tools used by the legal profession. Much has been written about the false citations and copyright claims arising from the use of AI-generated content in the legal realm, but little has been written about the way implicit bias shows up in the AI tools we're learning to love.

It's time to rectify that oversight. But first, a true confession: I used AI for my initial research into this subject. And why not? It's easy, convenient and more cost-effective than paying someone to do the work. As we all know, however, it should never be relied upon as the final word on a subject. My findings and conclusions — for better or worse — are ultimately my own.

Laws dealing with implicit AI bias

Effective Jan. 1, 2025, California enacted two laws dealing with AI bias. Government Code Section 11546.45.5 requires the state's Department of Technology, along with other interagency bodies, annually to conduct "a comprehensive inventory of all high-risk automated decision systems that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, any state agency." Among the inventory items are "[t]he measures in place, if any,

to mitigate the risks, including cybersecurity risk and the risk of inaccurate, unfairly discriminatory, or biased decisions, of the automated decision system."

Government Code Section 11549.63 finds that "because humans have explicit and implicit biases built into society, generative AI has the capacity to amplify these biases as it learns from input data, and that it is imperative to consider the implications for Californians of different races, ethnicities, genders, ages, and other characteristics in all AI inputs, outputs, and products." The law says that "no individual or group should be discriminated against on the basis of race, gender, age, religion, sexual orientation, or any other protected characteristic in the design, development, deployment, or use of AI systems."

At the federal level, Executive Order 14110, issued in October 2023, directed federal agencies to address AI-driven discrimination in hiring, housing, healthcare and other sectors. The secretary of labor was required to publish guidance for federal contractors regarding nondiscrimination in hiring involving AI and technology-based hiring systems; the Federal Housing Finance Agency and the Consumer Financial Protection Bureau were called on to ensure that regulated entities evaluated underwriting models for bias affecting protected groups and assessed automated processes for bias. That order was revoked in January 2025.

Cases looking at implicit AI bias

California and federal courts have increasingly confronted claims that artificial intelligence and algorithmic decision-making systems embed or amplify implicit bias, producing discriminatory outcomes in employment, housing, lending, insurance and criminal justice.

In *Mobley v. Workday, Inc.*, 740 F. Supp.3d 796 (2024), the Northern District of California ruled that an AI-powered hiring platform can be liable as an agent under Title VII, the ADEA, and the ADA for disparate impact discrimination, even without proof of intentional bias. The court articulated a principle of broad significance: drawing an artificial distinction between software decision-makers and human decision-makers would potentially gut anti-discrimination laws in the modern era. The court reasoned that nothing in the language of the federal anti-discrimination statutes distinguishes between delegating employment functions to an automated agent versus a live human one, and that employers could otherwise circumvent civil rights law by outsourcing discriminatory decisions to third-party AI tools.

California courts have applied the state's Racial Justice Act to address implicit bias in the criminal justice system Penal Code Section 745, *Jackson v. Superior Court*, 109 Cal.App. 5th 372 (2025). The legal landscape is rapidly evolving, with courts and regulators grappling with how existing civil rights frameworks apply to opaque, machine-learning-based systems.

AI in legal research

We've all read about the attorneys sanctioned for filing briefs with fabricated citations and hallucinations. We understand that attorneys have an obligation to verify AI-generated research results. (See *In re Martin*, 670 B.R. 636 (2025) (Bankr. N.D. Ill. 2025); *McCarthy v. DEA*, No. 24-274 (3d Cir. Mar. 27, 2026).) But fabrications and hallucinations are just the tip of the iceberg for the legal profession.

Attorneys must be able to answer a simple question: "Do you understand the software you're using?" Whether tracking prior cases, counseling clients or challenging legal orthodoxy, attorneys are increasingly using — and relying upon — AI tools. These tools are barely out of their infancy, and they are being developed seemingly at the speed of light. What we don't know about AI is probably as much as we do know.

AI can provide an unprecedented level of support; it is far more efficient and economical than traditional legal support. The AI tools commonly used by attorneys include Westlaw's Co-Counsel, Lexis-Nexis, Harvey, ChatGPT and Claude. The AI tools of Westlaw and Lexis-Nexis are closed, secure, legal universe tools, meaning these tools do not reach out into the internet to answer questions or prompts. These AI systems rely solely on the legal authorities contained in their legal software. Harvey is not closed, but it relies mostly on a closed universe and can break out to retrieve additional data. ChatGPT and Claude are not closed legal universes.

Bias in legal AI

AI did not start from a clean slate. All generative AI tools are trained using historic internet data, data into which implicit biases are embedded because their creators are human with human implicit biases. Legal AI is fine-tuned to the legal domain, but it does not "know" the law; it does not "understand" the law. It is merely performing pattern continuation. The 9th Circuit in *Malkeet Lnu v. Blanche*, — F.4th — (2026) acknowledged that generative legal AI is prone to "hallucinations"; it may cite real authorities but provide legally or factually incorrect information. Even advanced legal-specific AI models, the court said, are unlikely to fully solve the hallucination problem due to the complexity of the common law system.

Biases involving race, gender, sexual orientation, religion and other protected categories are inescapable in AI tools, because humans created the underlying data. In *Huskey v. State Farm Fire & Casualty Company*, No. 22 C 7014 (N.D. Ill 2023), a case involving AI used by insurance companies, the court noted that AI algorithms trained on historically biased data can create self-reinforcing cycles of discrimination. "Even when users do not input data about race, for example, algorithms can learn to combine other inputs correlated with race to produce discriminatory effects."

AI can also confirm incorrect user assumptions, based on the tendency of users to search for information that supports what they believe. This "confirmation bias" causes the LLM to ignore contradictory data, and users are prone to over-rely on its recommendations without doing further research. In the case of *In re Bryant*, 676 B.R. 352 (Bankr. M.D.N.C. 2025), the bankruptcy court cited research showing that even skilled psychologists were more likely to accept AI recommendations that confirmed their preliminary diagnoses.

Whether legal or non-legal AI, it is a summary of data that humans correct. Those "reviewing" humans have implicit biases, which are then transferred back to the generative AI at the time the humans input their "corrections." Just as is the case for insurance tools, legal tools will have inherent and implicit biases. But because the law is intended to be blind, treating everyone equally, this should raise huge red flags for the legal profession.

Counteracting AI bias

The first step in counteracting implicit bias in AI tools is recognizing that such bias exists. Attorneys must understand the limits of each AI tool they use and take affirmative steps to identify and address implicit biases.

The developers of generative AI are working to increase the level of trust in their systems, but the challenges are many and they may be intractable. Generative AI is a deep-learning transformer, a next token prediction, and bias is inherent in the system. An LLM will predict the next word or phrase from patterns in the training data it receives. Generative AI may be good at generating plausible text, but it predicts the most likely text. It struggles with rare, specialized ideas, and its predictions are not always correct. The answers may sound plausible, but they are not always defensible.

AI will actually suppress what it knows when placed under authority pressure. For example, AI can be convinced that $2+2=5$ through conversational pressure, social framing or prompt context. An attorney's pressure may thus insert bias into AI.

For this reason, it is incumbent on practitioners to take active steps to identify and address inherent bias in AI tools. At a minimum, attorneys should do the following:

Treat AI output no differently than the work product of junior associates;

Audit where the data came from, whether from ChatGPT or specialized legal AI software with a closed universe;

Only use vendors designed to keep practitioners accountable (legal AI software with a closed universe); and

Recognize their own internal biases (often implicit) and conduct themselves accordingly in using AI. For example, they should be careful not to use prompts that would encourage a biased result.

Conclusion

AI has opened a brave new world for legal practitioners. It provides stunningly quick and detailed answers for a wide range of questions. But, like Circe in the Odyssey, it may be a temptress that leads to doom. Attorneys who rely exclusively or primarily on AI tools could end up violating ethical obligations to their clients, the legal system and their own professional standards.

Attorneys who use AI must do so with their eyes wide open, understanding that, just as humans are flawed, so too is the technology that humans have created.

An attorney's work is just beginning whenever AI has been used.

Hon. Susan L. Greenberg (Ret.) *is a neutral with Signature Resolution who served 26 years as a San Mateo County Superior Court — 14 as a Commissioner and 12 as a Judge. Before her election to the bench, Judge Greenberg served as a Deputy District Attorney in San Mateo County, conducting more than 30 jury trials as a prosecutor before transitioning to private civil practice, where she handled personal injury, breach of contract, family law, real estate, and probate matters and mediated family law disputes.*