AI'S DOUBLE-EDGED PROMISE: EXAMINING ETHICS, BIAS, AND BUSINESS OUTCOMES IN AN AGE OF RAPID INNOVATION

Matthew J. Hitchcock*^

Synopsis: Artificial Intelligence (AI),[1] once relegated to science fiction and theoretical computer science, has evolved into a tool that has reshaped technology,[2] has set the business world on a new trajectory,[3] and could bring about the most significant redistribution of power in history.[4] Advancements in machine learning, neural networks, and data processing have propelled AI into practical, everyday applications[5] that impact critical sectors

 Matt Hitchcock is a member of The Florida Bar practicing in St. Petersburg, Florida. He earned his Juris Doctor and Master of Business Administration degrees from Stetson University in 2007 and his Bachelor of Business Administration from the same institution in 2004. Mr. Hitchcock holds a CIPP/US privacy certification and has been serving as in-house counsel since 2008. Over the past decade, he has gained extensive experience in the insurance industry, with a focus on the intersection of regulatory compliance and technology innovation. The views expressed in this Article are solely those of the author. This Article is dedicated to the memory of his mother, whose unwavering support and encouragement were instrumental in helping Matt achieve his dream of becoming an attorney.

^ © This Article was prepared with the collaborative assistance of artificial intelligence tools such as ChatGPT, Grammarly, Claude, and Perplexity, which were used for tasks such as research, outline development, drafting content, editing, and revisions. As the author, I take full responsibility for the content, including the accuracy of legal analysis and citations, which I have reviewed and verified to ensure compliance with scholarly standards and the ethical obligations of authorship.

1. In this Article, I use the term "artificial intelligence" or "AI" broadly to refer to a range of technologies capable of performing tasks that typically require human intelligence, including but not limited to machine learning, large language models (LLMs), and generative AI. While distinctions exist among these technologies, AI is used here as an umbrella term to focus on their legal and ethical implications rather than technical taxonomy.

2. Neha Soni et al., Impact of Artificial Intelligence on Businesses: From Research, Innovation, Market Deployment to Future Shifts in Business Models 2 (unpublished manuscript) (on file with the *Stetson Business Law Review*) (discussing AI's transformative effect on business models and innovation).

3. *Id.* at 11.

4. Mustafa Suleyman, *How the AI Revolution Will Reshape the World*, TIME (Sep. 1, 2023, at 7:05 EDT), https://time.com/6310115/ai-revolution-reshape-the-world/ (on file with the *Stetson Business Law Review*) (arguing that AI could prompt one of the most significant redistributions of power in modern history).

5. Soni et al., *supra* note 2, at 6.

such as finance, healthcare, and law.[6] AI-driven tools now manage investment portfolios, underwrite insurance policies, automate legal document reviews, diagnose medical conditions, and personalize educational experiences, demonstrating unprecedented efficiency and predictive accuracy.[7] Despite its enormous potential, AI's rapid integration into major areas of society raises significant ethical, legal, and human rights concerns.[8] We stand at a crucial juncture where AI offers substantial benefits, such as increased productivity, cost savings, and innovative problem-solving, while also introducing complex challenges that require our attention and scrutiny. Some of the most pressing issues involve the perpetuation of bias, lack of transparency, privacy concerns, and serious questions about accountability in AI decision-making. The breakneck pace at which AI is presently being adopted far outpaces the current regulatory environment, creating substantial gaps in oversight and ethical governance.[9] Attorneys are well-positioned to bridge these gaps and have an opportunity to shepherd their clients into an AI Age that is fair, transparent, and beneficial for all.

This article will explore the critical ethical issues associated with the use of AI in business, focusing particularly on the responsibilities of legal professionals and the potential for attorneys to be leaders during a pivotal moment in AI's evolution. At this juncture, it is evident that attorneys cannot avoid AI. Instead, they must navigate this challenging intersection of competence, candor to the tribunal, confidentiality, and accountability for AI-generated outcomes with the knowledge and skill required of attorneys. Additionally, this article will address AI bias, distinguishing between human-induced and algorithmic biases, and explore how well-designed AI systems can reduce existing bias while simultaneously improving business outcomes. This article aims to provide guidance for legal professionals,

6. *See generally* Zhiyu Zoey Chen et al., *A Survey on Large Language Models for Critical Societal Domains: Finance, Healthcare, and Law*, TRANSACTIONS ON MACH. LEARNING RSCH., Nov. 2024, at 1, 1 (exploring the applications of AI within finance, healthcare and law).

7. *See generally* STAN. INST. FOR HUM.-CENTERED A.I., AI INDEX REPORT 2025 (2025) (providing a comprehensive survey of global trends, applications, and impacts of artificial intelligence in various industries).

8. Yuzhou Qian, Keng L. Siau & Fiona F. Nah, *Societal Impacts of Artificial Intelligence: Ethical, Legal, and Governance Issues*, 3 SOCIETAL IMPACTS, 2024, at 1, 2.

9. Chen et al., *supra* note 6, at 30.

business leaders, and policymakers through practical case studies, analysis of the regulatory environment, and strategic recommendations. Ultimately, it advocates for a proactive approach to ethical AI development, balancing technological innovation with responsible oversight.

TABLE OF CONTENTS

I. INTRODUCTION

AI's journey began in the 1950s, when early pioneers such as Alan Turing and John McCarthy laid the conceptual foundation.[10] The following decades witnessed multiple periods of rapid progress and enthusiasm (AI Summers), followed by stagnation and skepticism (AI Winters).[11] Over time, AI became a powerful tool for those with the necessary resources and knowledge.[12] The significant computing power, vast amounts of data, and specialized expertise needed to create AI served as an insurmountable barrier to entry for most, resulting in AI primarily being a tool of major corporations and well-funded research institutions.[13] Despite these limitations, over time, AI was integrated into everyday business operations through innovations such as automated customer

10. Selmer Bringsjord & Naveen Sundar Govindarajulu, *Artificial Intelligence*, STANFORD ENCYC. OF PHIL. ARCHIVE (July 12, 2018), https://plato.stanford.edu/archives/spr2025/entries/artificial-intelligence/ [https://perma.cc/83NA-SUJD].

11. *See* Amirhosein Toosi et al., *A Brief History of AI: How to Prevent Another Winter (A Critical Review)*, 16 PET CLINICS 449, 459 (2021).

12. *Id*. at 452–57 (discussing how limited access to computing resources and institutional support constrained early AI development and limited its reach outside of elite institutions).

13. *Id.*

service chatbots,[14] algorithm-driven stock trading,[15] and predictive analytics in marketing.[16]

During the 1980s and 1990s, AI remained behind the scenes, embedded in proprietary expert systems and enterprise software, rather than being directly accessible to the average professional or consumer;[17] now that dynamic has changed.[18] Thanks to cloud computing, open-source AI models, and user-friendly platforms, AI is no longer confined to the exclusive domain of tech giants and data scientists.[19] AI has become an integral part of our daily lives, often in ways we do not even recognize.[20] Every time we unlock our smartphones using facial recognition, receive personalized recommendations on streaming services, or dictate a message using voice-to-text, we engage with AI.[21] Virtual assistants like Siri and Alexa, real-time language translation tools, and automated email sorting in our inboxes demonstrate how AI has quietly woven itself into the fabric of modern life.[22]

In today's business world, access to AI has grown exponentially, allowing companies of all sizes to leverage its

14. *See* Albérico Travassos Rosário & Ricardo Jorge Raimundo, *The Integration of AI and IoT in Marketing: A Systematic Literature Review*, 14 ELEC. 1854, 1868 (2025) (discussing the role of AI-driven chatbots and automation in transforming customer engagement).

15. *See generally Artificial Intelligence in Capital Markets: Use Cases, Risks and Challenges*, INT'L ORG. OF SEC. COMM'NS (2025), https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf [https://perma.cc/T2G3-5H5M] [hereinafter INT'L ORG. OF SEC. COMM'NS] (analyzing the use of AI in algorithmic trading and investment decision-making).

16. *See generally* Berk Yilmaz & Huthaifa I. Ashqar, *Towards Equitable AI: Detecting Bias in Using Large Language Models for Marketing*, ARXIV (Feb. 18, 2025) (unpublished manuscript) (on file with the *Stetson Business Law Review*) (exploring the application of AI and predictive analytics in modern marketing practices).

17. Toosi et al.*, supra* note 11, at 457.

18. Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base – Analyst Note*, REUTERS (Feb. 2, 2023, at 10:33 EST), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ [https://perma.cc/QA2K-Q266] (reporting on ChatGPT's monumental growth as a consumer application).

19. Alice Gomstyn & Alexandra Jonker, *Democratizing AI: What Does it Mean and How Does it Work?,* IBM (Nov. 5, 2024), https://www.ibm.com/think/insights/democratizing-ai [https://perma.cc/5L39-L5VP] (highlighting efforts to make AI accessible to a broader range of users beyond machine learning experts).

20. Bernard Marr, *The 10 Best Examples of How AI Is Already Used in Our Everyday Life*, FORBES (Dec. 10, 2021, at 8:30 ET), https://www.forbes.com/sites/bernardmarr/2019/12/16/the-10-best-examples-of-how-ai-is-already-used-in-our-everyday-life/ [https://perma.cc/D2WZ-PZ5W] (providing examples of AI in daily activities including unlocking phones with face ID, using digital voice assistants, and receiving personalized content recommendations).

21. *Id.*

22. *Id.*

capabilities.[23] Small businesses can now access AI-powered tools to manage customer relations, automate bookkeeping, and enhance marketing strategies.[24] In legal practice, AI-driven contract analysis, legal research, and e-discovery tools are streamlining workflows, saving time, and reducing costs.[25] The democratization of AI has begun, and with it comes the challenge of ensuring it is used in an ethical, responsible, and fair manner.[26] This moment is critical because AI is still in its early stages of widespread adoption, and the regulatory and ethical frameworks surrounding it remain unsettled.[27] Decisions made by businesses, policymakers, and legal professionals today may determine whether AI creates a better or worse tomorrow.[28] Without proactive governance, AI could entrench biases and create new liabilities for organizations.[29] However, with thoughtful oversight and implementation, AI has the potential to enhance access to justice, improve business outcomes, and drive innovation in ways that benefit society as a whole.[30] Establishing responsible AI governance now is not just an opportunity; it's an imperative.

It should be acknowledged that some people view the current AI zeitgeist with derision, believing it to be overhyped and ultimately harmful.[31] While these views have merit, it is undeniable that AI is currently having a significant impact on

23. *See generally* Fadeke Adegbuyi, *AI for Small Business: Applications, Benefits, and Risks*, SHOPIFY (May 8, 2025), https://www.shopify.com/blog/ai-for-small-business [https://perma.cc/3SM5-BRTN] (discussing how AI tools assist small businesses in product development, marketing, and sales).

24. *Id.*

25. *See* Chukwuemezie Charles Emejuo et al., *The Impact of Artificial Intelligence on Legal Practice: Enhancing Legal Research, Contract Analysis, and Predictive Justice*, 14 INT'L J. SCI. & RSCH. ARCHIVE 603, 604 (2025) (discussing how AI technologies are transforming legal research, contract analysis, and predictive justice by improving efficiency and accuracy).

26. Gomstyn & Jonker, *supra* note 19.

27. Qian, Siau & Nah, *supra* note 8, at 2.

28. *Id.*

29. Tim Mucci & Cole Stryker, *What Is AI Governance?*, IBM (Oct. 10, 2024), https://www.ibm.com/think/topics/ai-governance [https://perma.cc/959H-BWXN].

30. Enas Mohamed Ali Quteishat et al., *Exploring the Role of AI in Modern Legal Practice: Opportunities, Challenges, and Ethical Implications*, 20 J. ELEC. SYS. 3040, 3040–41 (2024).

31. David Widder & Mar Hicks, Watching the Generative AI Hype Bubble Deflate (Aug. 16, 2024) (unpublished manuscript) (on file with the *Stetson Business Law Review*) (arguing that the hype surrounding generative AI is waning and cautioning that its overstated promises have already caused lasting social and ethical harm).

many areas of society, including the practice of law.[32] Because attorneys are held to a high standard of professionalism, many may be justifiably cautious about adopting AI into their legal practice. However, as the use of AI spreads, attorneys must develop a reasonable level of knowledge and skill concerning the benefits and risks of AI. Doing so will ensure they uphold their professional obligations to their clients. The following part will examine the ethical and legal implications of the growing presence of AI in the legal profession and explore the role of attorneys in shaping its responsible deployment.

II. THE ATTORNEY'S ETHICAL OBLIGATIONS WITH AI

For some observers, the pace at which AI has moved from the margins to the mainstream is a testament to its great potential.[33] As the previous part outlined, AI is reshaping industries and introducing new capabilities into business and everyday life, and the legal profession is no exception. The market for legal AI tools is expected to reach $1.75 billion in 2025 and more than double to $3.9 billion by 2030,[34] a strong indication that AI has arrived in the legal field and is expected to grow rapidly over the next few years. As AI takes a larger role in legal work, attorneys must ensure their use of this tool is consistent with their professional obligations. Chief among these are the duties of technological competence, candor to the tribunal, maintaining client confidence, and responsibility for AI-generated work product.

A. AI and the Technology Competence Rule

Model Rule 1.1 of the ABA Model Rules of Professional Conduct requires attorneys to provide competent legal

32. Benjamin Alarie, Anthony Niblett & Albert Yoon, *How Artificial Intelligence Will Affect the Practice of Law,* 68 U. TORONTO L.J. 106, 116 (2018) (acknowledging skepticism about AI hype but arguing that AI is already reshaping the practice of law in substantive ways, including legal research, prediction, and client services).

33. Adam Blandin, Alexander Bick & David Deming, *The Rapid Adoption of Generative AI* (Fed. Rsrv. Bank of St. Louis, Working Paper No. 2024-027A, 2024), (finding that almost 40 percent of U.S. adults used generative AI two years post-launch, twice the rate of the internet's two-year adoption).

34. *Legal AI Market Size, Share & Trends Analysis Report by Component, by Technology, by Application, by End-use, by Region, and Segment Forecasts, 2025–2030*, GRAND VIEW RSCH. (2024), https://www.grandviewresearch.com/industry-analysis/legal-ai-market-report [https://perma.cc/4728-EGUA] (summarizing key findings at beginning of page).

representation to clients.[35] In 2012, the American Bar Association (ABA) amended Comment 8 to Rule 1.1 to clarify that this includes competence with current forms of technology.[36] Comment 8 to Rule 1.1 now states, "[t]o maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology."[37] In today's legal environment, relevant technology undoubtedly includes AI.

The benefits of AI for busy legal professionals are evident. The ability to quickly generate legal documents, analyze case law, and predict litigation outcomes presents tremendous opportunities for efficiency.[38] However, an attorney who accepts AI-generated legal research or analysis without verifying its accuracy may produce flawed work and violate their ethical duty of competence.[39] This is demonstrated in recent cases where legal professionals misused AI tools in litigation.[40] After learning about other legal professionals sanctioned and disciplined for the improper use of AI, many attorneys are understandably hesitant to employ AI tools in their practice. However, avoidance is not a solution. Attorneys must understand AI, even if they do not intend to use it in their legal work. Recent misuse of AI in litigation has shown that attorneys need to be more vigilant and prepared to identify any improper or unethical use of AI.[41] Thus, whether an attorney plans to use it or not, they must acquire a basic understanding of the benefits and risks associated with AI in order to fulfill their professional duties in a world rich with AI tools.

Another critical challenge is the "black box" nature of many AI models.[42] This issue stems from their complexity and lack of transparency.[43] Unlike traditional research, where attorneys

35. *See* MODEL RULES OF PRO. CONDUCT r. 1.1 (A.B.A. 2023).

36. *Id.* at cmt. 8.

37. *Id.*

38. *See generally* Emejuo et al., *supra* note 25 (stating the use of AI in predictive justice and contract analysis).

39. A.B.A. Standing Comm. on Ethics & Prof. Resp., Formal Op. 512, at 4 (2024) [hereinafter A.B.A. Formal Op. 512].

40. *See* Damien Charlotin, *AI Hallucination Cases* (2025), https://www.damien charlotin.com/hallucinations/ [https://perma.cc/B2LS-RAJY] (providing a running list of litigation involving the misuse of generative AI and hallucinated case citations).

41. Lisa Z. Rosenof, *The Fate of Comment 8: Analyzing a Lawyer's Ethical Obligation of Technological Competence*, 90 U. CIN. L. REV. 1321, 1339 (2022).

42. Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 891–92 (2018).

43. *Id.*

follow citations and reasoning, AI-driven tools often provide results without clear explanations of how the AI reached the conclusions it did.[44] This opacity makes assessing reliability difficult, increasing the risk of errors or misleading conclusions.[45] Thus, to provide competent representation, lawyers must understand how AI tools function, their limitations, and how improper data inputs and training methods may result in biased AI outputs that may be harmful to clients.[46]

The duty of competence also extends to advising clients on AI-related legal risks.[47] Many businesses are rapidly integrating AI into decision-making, from automated business processes to contract negotiations and risk assessments.[48] These organizations may face legal consequences if AI systems inadvertently discriminate, breach contractual obligations, or create regulatory exposure.[49] Attorneys must understand these risks and be prepared to counsel clients on compliance strategies, risk mitigation, and the evolving legal landscape surrounding AI.

Meeting the duty of competence requires a reasonable degree of knowledge and skill concerning current technologies and their associated benefits and risks.[50] To uphold the duty of technical competence, attorneys should: stay informed about advancements in AI and their legal implications through continuing legal education (CLE) programs, professional associations, and industry publications;[51] evaluate AI tools critically, ensuring they meet standards of accuracy, reliability, and ethical integrity before integrating them into legal practice;[52] understand AI biases and limitations, recognizing that AI outputs can reflect and amplify

44. *Id.* at 901.
45. *Id.*
46. N.Y.C. Bar Ass'n, Comm. on Prof. Ethics, Formal Op. 2024-5 (2024).
47. A.B.A. Formal Op. 512, *supra* note 39, at 9 (discussing competence and Model Rule 1.1).
48. Alex Singla et al., *The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value,* MCKINSEY & CO. (May 30, 2024), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024 [https://perma.cc/Z25R-4L2X].
49. INT'L ORG. OF SEC. COMM'NS, *supra* note 15, at 40, 47, 50 (discussing the risks companies face when deploying AI in financial markets).
50. *See* MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (A.B.A. 2018) (emphasizing the duty of technological competence).
51. A.B.A. Formal Op. 512, *supra* note 39, at 10.
52. *Id.* at 4.

systemic biases in training data;[53] and educate clients on AI-related risks.[54]

With AI's growing role in law and society, attorneys must remain focused on their professional responsibility of technological competence.[55] By adhering to the above recommendations, attorneys should develop the knowledge and the skills necessary to act as trusted advisors to clients as they navigate the complexities of AI adoption. Furthermore, attorneys who embrace their responsibility and seek to understand and critically engage with AI will not only fulfill their professional obligations but will also be well-positioned to help shape the development of ethical AI systems.

B. Candor to the Tribunal: The Risks of AI Misuse in Litigation

Competence is a critical first step because attorneys who fulfill their duty of competence should be able to recognize and avoid other issues that can arise from the improper use of AI, such as the duty of candor to the tribunal. Model Rule 3.3 requires lawyers to act with candor toward the tribunal and prohibits them from knowingly making false statements of fact or law or failing to correct false statements previously made to the court.[56] Nevertheless, attorneys are making headlines due to legal filings that contain fake case citations.[57] This issue arises from the fact that AI is prone to a phenomenon known as 'Hallucination,' where the AI generates text that appears plausible but is absolute fiction.[58] Attorneys who fail to verify AI-generated citations may find themselves not only embarrassed but also subject to sanctions,

53. Emejuo et al., s*upra* note 25, at 607.

54. A.B.A. Formal Op. 512, *supra* note 39, at 8.

55. *Id*. at 8–9.

56. MODEL RULES OF PRO. CONDUCT r. 3.3 (A.B.A. 2018).

57. Debra Cassens Weiss, *Sanctions Imposed for 'Collective Debacle' Involving AI Hallucinations and 2 Law Firms, Including K&L Gates*, ABAJOURNAL (May 14, 2025, at 12:50 CDT), https://www.abajournal.com/web/article/judge-imposes-sanctions-for-collective-debacle-involving-ai-hallucinations-and-2-law-firms-including-k (on file with the *Stetson Business Law Review*) (Despite using commercial AI tools designed to mitigate hallucinations, attorneys from two law firms were sanctioned after relying on AI-generated case citations that turned out to be fictitious. The court criticized the firms' inadequate oversight and called the incident a "collective debacle.").

58. *See* Charlotin, *supra* note 40.

as recent cases have demonstrated.[59] These incidents are rarely borne from malice, and most involve attorneys experimenting with new tools, facing time pressure, or misunderstanding the technology's limitations.[60] Nevertheless, the attorney is ultimately responsible for AI-hallucinated mistakes.[61]

C. Confidentiality Concerns: Hidden Risks in Everyday Use

While courtroom missteps get headlines, more subtle risks to client confidentiality often go unnoticed but can be just as dangerous. Under Model Rule 1.6, attorneys are prohibited from revealing "information relating to the representation of a client unless the client gives informed consent" or the disclosure is otherwise permitted.[62] Using AI tools, especially those accessed via cloud platforms or browser-based applications, can inadvertently expose sensitive client information.[63] When an attorney pastes portions of a draft complaint, contract, or due diligence materials into an AI platform, that data may be stored, processed, or even used to train the model, depending on the terms of service and the user's settings.[64]

---

59. *See, e.g.*, Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 466 (S.D.N.Y. 2023) (sanctioning attorneys for submitting a brief with fictitious case citations generated by ChatGPT); Park v. Kim, 91 F.4th 610, 614–16 (2d Cir. 2023) (referring attorney for potential discipline for including fake, AI-generated legal citations in a filing); Kruse v. Karlen, 692 S.W.3d 43, 53 (Mo. Ct. App. 2024) (dismissing appeal because litigant filed a brief with multiple fake, AI-generated legal citations); Gauthier v. Goodyear Tire & Rubber Co., No. 1:23-CV-00281 (E.D. Tex. Nov. 25, 2024) (imposing $2,000 sanction and mandatory AI-related CLE for submitting AI-generated fictitious citations); Al-Hamim v. Star Hearthstone, LLC, 2024 COA 128, ¶ 4 (Colo. App. 2024) (declining to sanction pro se litigant but warning that future filings with AI-generated hallucinations may result in sanctions); Concord Music Grp., Inc. v. Anthropic PBC, No. 5:24-cv-03811 (N.D. Cal. May 15, 2025) (attorney admitted responsibility for incorrect citation in expert report caused by AI hallucination); Kohls v. Ellison, No. 0:24-cv-00123 (D. Minn. Jan. 25, 2025) (excluding expert declaration based on AI-generated hallucinated citations).

60. *See* Jack Newsham, *AI Hallucination in Court Documents Are a Growing Problem, and Data Shows Lawyers Are Responsible for Many of the Errors,* BUS. INSIDER (May 27, 2025, at 6:33 ET), https://www.businessinsider.com/increasing-ai-hallucinations-fake-citations-court-records-data-2025-5 (on file with the *Stetson Business Law Review*); *see also* James O'Donnell, *How AI Is Introducing Errors into Courtrooms,* MIT TECH. REV. (May 20, 2025) https://www.technologyreview.com/2025/05/20/1116823/how-ai-is-introducing-errors-into-courtrooms/ [https://perma.cc/N7VJ-4HMC].

61. *See* Newsham, *supra* note 60.

62. MODEL RULES OF PRO. CONDUCT r. 1.6 (A.B.A. 2018).

63. Nicholas Daniel Seger, *Understanding the Risks of Uploading Client Information to Generative AI Platforms*, A.B.A. (Jan. 16, 2024), https://www.americanbar.org/groups/young_lawyers/resources/tyl/practice-management/risks-uploading-client-information-generative-ai-platforms/ (on file with the *Stetson Business Law Review*).

64. *Id.*

Therefore, before implementing any new AI tools, an attorney should consider:

(1) Where is the data that is entered into the platform going? Is the data stored in another country? If so, is that an issue?

(2) Who can access the data that is entered into the platform? Does the platform vendor have a trustworthy reputation?

(3) Is the data encrypted at rest and in transit?

(4) Does the platform provider have data retention and deletion policies?

(5) Does the platform vendor have a SOC 2 report[65] or a comparable assurance document demonstrating that it has implemented effective controls to protect customer data?

(6) Do you need a data processing agreement (DPA) when dealing with international clients?

Failing to consider these questions may result in the attorney violating the duty of confidentiality, even though the attorney never intended to expose the client's information.[66] Also, consider the use of AI in litigation management platforms, e-discovery systems, transcription software, or email filtering tools. While many of these are essential to modern practice, their use must be paired with a basic understanding of how client information is managed. Attorneys must always vet vendors carefully, ensure they are bound by appropriate contractual obligations, train staff on the proper use of technology, and adequately supervise those who provide legal assistance, whether human or an algorithm.[67]

65. A SOC 2 (System and Organization Controls 2) report is an independent audit report that evaluates a service organization's security, availability, processing integrity, confidentiality, and privacy controls. *See SOC 2 – SOC for Service Organizations: Trust Services Criteria*, AICPA & CIMA, https://www.aicpa-cima.com/topic/audit-assurance/audit-and-assurance-greater-than-soc-2 [https://perma.cc/8P2K-G8V7] (last visited Sep. 20, 2025).

66. *See* MODEL RULES OF PRO. CONDUCT r. 1.6(c), cmt. 18 (A.B.A. 2018) (requiring lawyers to "make reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to, information relating to the representation of a client," including when using "technology and devices" and "third-party service providers").

67. A.B.A. Standing Comm. on Ethics & Pro. Resp., Formal Op. 477R 9–10 (2017) [hereinafter A.B.A. Formal Op. 477R].

D. Responsibility for AI-Generated Outcomes

The 2012 amendments to the ABA Model Rules also contained a very small change to the title of Model Rule 5.3 that greatly expanded the scope of the rule. The change is a very subtle but purposeful change from "Responsibilities Regarding Nonlawyer Assistants" to "Responsibilities Regarding Nonlawyer Assistance."[68] This change from "assistant" to "assistance" clarified that, just as attorneys cannot delegate their professional duties to a human assistant, they also cannot delegate their professional responsibility to an AI assistant.[69] No matter how advanced or efficient an AI tool may be, its use does not relieve an attorney of their professional obligations. Regardless of whether it is an attorney, a human legal assistant, or an AI tool that drafts a brief, revises a contract, or summarizes discovery, the attorney remains fully responsible for the final work product.[70]

AI should augment the role of people in the workplace, not replace them.[71] Overreliance on these tools risks circumventing the diligence, contextual understanding, and professional scrutiny the legal profession demands.[72] Attorneys must ensure that AI tools are carefully selected, appropriately deployed into legal workflows, and adequately supervised.[73] Consider the following examples of well-intentioned use of AI that can lead to ethical pitfalls. An attorney might rely on AI to summarize a deposition transcript but fail to notice that the summary distorts a key admission, potentially affecting case strategy.[74] In another instance, a contract clause generated by an AI tool may introduce indemnification language that is detrimental to the client's position, escaping the attorney's notice until after execution.[75] Similarly, during AI-assisted document review, an attorney who does not configure appropriate filters may inadvertently allow the

68. *ABA Ethics Rules and Generative AI,* THOMPSON REUTERS LEGAL BLOG (Mar. 27, 2025), https://legal.thomsonreuters.com/blog/generative-ai-and-aba-ethics-rules/ [https://perma.cc/N6CD-DKXY] (discussing the amendment to Model Rule 5.3 changing "assistant" to "assistance," thereby extending supervision duties to non-human actors such as AI).

69. *Id.*

70. A.B.A. Formal Op. 512, *supra* note 39, at 2–4.

71. Tim O'Reilly, *AI First Puts Humans First*, O'REILLY RADAR (May 28, 2025), https://www.oreilly.com/radar/ai-first-puts-humans-first/ [https://perma.cc/Q9ZB-4CPA].

72. A.B.A. Formal Op. 512, *supra* note 39, at 3–4.

73. *Id.*

74. *Id.*; *see also* A.B.A. Formal Op. 477R, *supra* note 67, at 2–3.

75. A.B.A. Formal Op. 512, *supra* note 39, at 3.

disclosure of privileged materials, exposing the client to unnecessary risk.[76] These are not hypothetical outliers. They demonstrate how easily a lapse in oversight may violate the duty of competence and the obligation to protect client confidences.[77] Ultimately, lawyers must understand the capabilities and limitations of AI tools and take meaningful steps to supervise their use.[78] Responsibility must always rest with the human actors who design, deploy, and use AI, because accountability for harm caused by it cannot be shifted to AI. The duty of competence, candor, diligence, and confidentiality must always remain with the attorney.

III. AI AND THE PERPETUATION OR ELIMINATION OF BIAS

Just as the rules of professionalism require attorneys to remain accountable for the outputs of AI tools, the duty of competence demands that lawyers understand the broader ethical implications of AI systems, particularly the risks of bias and the harm that can result from biased or otherwise flawed AI outputs.[79] As AI reshapes decision-making across government, business, and the law, it introduces powerful capabilities, but it also creates serious challenges for society.[80] Perhaps the most significant among these challenges is the concern that AI systems can perpetuate or amplify historical biases embedded in the data on which they are trained.[81] However, if designed and governed effectively, AI has the potential to be a tool for identifying and mitigating bias while simultaneously improving fairness, efficiency, and business performance.[82] Attorneys who advise clients on the legal risks, ethical responsibilities, and governance structures surrounding AI design and deployment must recognize that AI has the potential for substantial benefit and significant

76. *Id.* at 6; A.B.A. Formal Op. 477R, *supra* note 67, at 3.

77. *See* MODEL RULES OF PRO. CONDUCT r. 1.1, 1.6 (A.B.A. 2018); *see also* A.B.A. Formal Op. 512, *supra* note 39, at 3–4.

78. A.B.A. Formal Op. 512, *supra* note 39, at 3–4.

79. *Id.*

80. Qian, Siau & Nah, *supra* note 8, at 2.

81. *See* Emilio Ferrara, *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies*, 6 SCI. 3, 4 (2023) (exploring how AI systems can both amplify bias when trained on flawed data and serve as tools to identify and reduce bias when properly designed and implemented).

82. *Id.* at 7.

harm.[83] Thus, obtaining a basic level of understanding of AI bias is essential for developing the competence required of legal professionals.

A. Understanding AI Bias

To properly advise clients, attorneys must understand how bias may arise in AI systems. Such bias generally originates from the data on which the models are trained and the design choices made during development.[84] AI learns from historical patterns, and if those patterns reflect prejudice, injustice, or underrepresented sampling, the AI will reproduce them.[85] Bias can manifest in many forms, such as racial, gender, socioeconomic, and geographic, as well as in many areas of society, including hiring, lending, public safety, legal protections, and judicial sentencing.[86] The following case study provides an example of how bias can arise in the hiring process and the challenges of designing an AI system to overcome bias.

Case Study 1: Bias in Hiring Algorithms

In the mid-2010s, Amazon's internal AI team developed a tool designed to streamline the hiring of software engineers and other technical talent. The goal was to reduce manual resume review and improve efficiency in talent acquisition. The AI system was trained on a decade of hiring data, primarily consisting of resumes submitted to Amazon over the previous ten years. However, the dataset used to train the model reflected historical imbalances in the tech industry, particularly the underrepresentation of women in technical roles. As a result, the AI model began to associate success with male-dominated patterns and penalize signals correlated with female applicants.[87]

---

83. *See generally* Linda Pressly & Esperanza Escribano, *Police Algorithm Said Lina Was at "Medium" Risk. Then She Was Killed*, BBC NEWS (Apr. 19, 2025), https://www.bbc.com/news/articles/clyw7g4zxwzo [https://perma.cc/XG9V-NZDL] (reporting on the failure of a police risk assessment AI to flag a domestic violence victim as high risk, ultimately resulting in her death, and illustrating how well-intentioned AI systems can lead to tragic consequences).

84. Ferrara, *supra* note 81, at 6.

85. *Id.* at 2.

86. *Id.* at 4.

87. Jeffrey Dastin, *Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018, at 20:50 ET), https://www.reuters.com/

Even after Amazon's developers attempted to neutralize the bias by removing explicit gender markers, the system continued to infer gender through proxies. Contextual clues — such as participation in women's chess clubs or attendance at all-women's colleges — led the AI to downgrade certain applications.[88] These workarounds proved ineffective because the AI model had already internalized deeply embedded statistical correlations that reproduced discriminatory outcomes.[89] Ultimately, Amazon abandoned the tool in 2018 before it was ever deployed in live hiring decisions. While the company did not publicly release the AI model or its technical specifications, reports of the tool's performance sparked widespread debate about fairness, accountability, and bias in AI decision-making.

This example illustrates several critical legal and ethical concerns. First, it underscores the risk of bias in training data. The AI system was likely not designed to discriminate. However, because it learned from real-world data encoded with historical biases, it recognized and replicated those patterns.[90] Second, the case exemplifies proxy discrimination. This is because, despite removing the protected characteristics, such as gender, the model continued to use statistically correlated proxies to reach the same discriminatory outcomes. This underscores the limitations of deidentification as a bias mitigation strategy.[91] Third, the lack of transparency and explainability surrounding the tool raises questions about accountability. Without access to the model's architecture or rationale, it becomes difficult for external stakeholders, including regulators and courts, to evaluate whether the system complies with anti-discrimination laws.[92] Finally, this case highlights the legal risks and corporate responsibilities associated with AI-driven decision-making in employment. Had the tool been used in practice, Amazon could have faced liability

article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG [https://perma.cc/RT8Q-S9KJ].

88. *Id.*

89. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact,* 104 CALIF. L. REV. 671, 691–92 (2016) (discussing AI proxies and historical bias in training data that leads to discriminatory outcomes in automated systems).

90. *Id.*

91. *Id.*

92. Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation,"* 38 AI MAG. 50, 55 (2017).

under Title VII of the Civil Rights Act or similar anti-discrimination laws, even absent a discriminatory intent.[93]

Amazon's failed recruitment tool is a cautionary tale. It shows that even sophisticated, well-resourced organizations may inadvertently create systems that violate legal and ethical standards. This is why attorneys advising clients on the development and deployment of AI tools, especially in employment and other high-stakes contexts, must go beyond surface-level assessments of legal compliance. A thorough review of the AI system's full lifecycle, from training data selection to post-deployment monitoring, is essential to mitigate bias, ensure fairness, and uphold the rule of law.

B. Human vs. Machine Bias

Bias is certainly not unique to machines. Humans are also prone to implicit bias and systemic prejudice. However, AI systems introduce new complexity due to the scale, speed, and opacity of their decision-making processes.[94] While human decisions, however flawed, can generally be explained, contested, and held accountable, many AI systems function as "black boxes," relying on complex statistical models that may be difficult to interpret or justify, even for the AI's creators.[95]

Several key distinctions exist between human and machine bias. First, AI systems often operate with greater opacity.[96] Whereas human bias may be inferred through statements or conduct, AI bias can be embedded in data preprocessing, feature selection, or algorithmic weighting, which are not readily observable.[97] Second, AI systems can reinforce bias through feedback loops.[98] For example, a predictive policing model that

93. *See* 42 U.S.C. § 2000e-2(k) (2025) (providing that employment practices causing disparate impact on the basis of protected characteristics are unlawful under Title VII unless job-related and consistent with business necessity).

94. Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 898–99 (2018).

95. *Id.* at 891.

96. *Id.*

97. *See* Sandra Wachter et al., *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI,* 41 COMPUT. L. & SEC. REV. 105567, 105573–75 (2021) [https://doi.org/10.1016/j.clsr.2021.105567] (discussing the opaque nature of AI decision-making and challenges to legal transparency and accountability).

98. Nicolò Pagan et al., A Classification of Feedback Loops and Their Relation to Biases in Automated Decision-Making Systems 1–2 (May 10, 2023) (unpublished manuscript) (on file with the *Stetson Business Law Review*).

disproportionately sends officers to specific neighborhoods will generate data showing higher arrest rates in those areas, thereby validating and perpetuating its biased assumptions.[99] Third, AI operates at an unmatched scale and speed.[100] An AI decision-making tool can process thousands of actions per minute, amplifying the reach and impact of flawed outputs far beyond that of any human actor.[101]

Despite these issues, AI also introduces a degree of consistency and auditability that human decision-making lacks. A biased human may not recognize or admit they are biased, whereas with AI, patterns can be detected and addressed through tools such as model audits, input-output testing, and counterfactual analysis.[102] For attorneys, this presents a complex challenge of advising clients on avoiding overt discrimination and the subtle forms of disparate impact that may arise from opaque AI decision-making. As the following case study demonstrates, a lack of transparency in AI decision-making can raise significant questions around fairness and justice.

Case Study 2: Judicial Risk Assessment Tools – COMPAS

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) AI tool, developed by Northpointe (now Equivant), is an AI risk assessment system widely used in the United States to inform pretrial release, sentencing, and parole decisions.[103] Its primary purpose is to evaluate a defendant's likelihood of recidivism based on various inputs, such as criminal history, age, employment status, and responses to survey

99. Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice,* 94 N.Y.U. L. REV. ONLINE 192, 200-05 (2019) (explaining how predictive policing systems amplify bias through reinforcement of flawed data).

100. Carlos Batallas, *When AI Meets the Laws of War*, IE INSIGHTS (Oct. 3, 2024), https://www.ie.edu/insights/articles/when-ai-meets-the-laws-of-war/ [https://perma.cc/9B6N-73NM].

101. Finale Doshi-Velez & Mason Kortz, *Accountability of AI Under the Law: The Role of Explanation* (Berkman Klein Ctr. Working Paper No. 2017-6, 2017) (highlighting how AI systems differ from human decision-making in terms of scale, consistency, and the challenges of tracing responsibility).

102. Wachter et al., *supra* note 97, at 105573–75 (discussing the benefits of AI systems for enabling consistency, auditability, and post hoc analysis of bias, while warning that legal standards may still be difficult to operationalize).

103. *Practitioner's Guide to COMPAS Core*, EQUIVANT 1 (2017), https://cjdata.tooltrack.org/sites/default/files/2018-10/Practitioners_Guide_COMPASCore_121917.pdf [https://perma.cc/4ZZ6-DLM8].

questions.[104] However, its use has drawn significant scrutiny, particularly concerning racial bias and the lack of transparency in its methodology.[105]

A 2016 investigation by ProPublica analyzed the COMPAS scores of over 7,000 individuals arrested in Broward County, Florida, and compared the predictions to actual recidivism outcomes over two years.[106] The results showed that black defendants were nearly twice as likely as white defendants to be incorrectly classified as high risk (false positives), while white defendants were more often incorrectly classified as low risk (false negatives).[107] These disparities existed despite similar actual rates of reoffending between the two groups.[108]

A significant criticism of COMPAS is that it lacks transparency and operates as a black-box system.[109] The developers of COMPAS have refused to disclose the specific factors and weights used to generate risk scores because the system is proprietary.[110] The end result is that defendants are deprived of the opportunity to meaningfully rebut the evidence against them, even when the tool's output may influence judicial decisions about the defendant's liberty.[111] This lack of transparency raises serious due-process concerns.[112]

The COMPAS controversy exemplifies a broader concern that AI systems trained on historical data may encode and perpetuate societal biases. Suppose the data reflects patterns of over-policing or sentencing disparities rooted in systemic racism. In that case, the AI will learn to associate race-adjusted proxies with higher risk and replicate rather than correct past injustices. Moreover, the deployment of such tools in high-stakes domains such as criminal

---

104. *Id.*

105. *See* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/H59J-VFZK]. *But see* William Dieterich et al., *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, NORTHPOINTE INC. (July 8, 2016), https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html (on file with the *Stetson Business Law Review*) (critiquing ProPublica's methodology and arguing COMPAS performs equally across racial groups).

106. Angwin et al., *supra* note 105.

107. *Id.*

108. *Id.*

109. *Id.*

110. *Id.*

111. *See* State v. Loomis, 881 N.W.2d 749, 753 (Wis. 2016) (upholding use of COMPAS system but warning that lack of transparency may raise due process concerns).

112. *Id.*

sentencing can amplify the harm for people facing incarceration or release decisions.

Beyond civil liberties, this case also illustrates emerging areas of legal risk for governments and vendors that adopt or create biased AI tools. For example, the Texas Responsible Artificial Intelligence Governance Act (TRAIGA, H.B. 149) expressly bans AI systems that infringe, restrict, or impair any rights guaranteed under the U.S. Constitution or that unlawfully discriminate against protected characteristics such as race, color, sex, age, or disability.[113] The law empowers the Texas Attorney General to enforce these provisions with robust penalties, including civil fines ranging from $10,000 to $200,000 per violation and $2,000 to $40,000 per day for ongoing noncompliance.[114] Government agencies and private sector developers face mandatory disclosures and a 60-day cure period before more severe penalties apply.[115]

While discriminatory AI may expose public agencies to constitutional challenges under the Equal Protection Clause,[116] at the same time, developers and vendors may face tort, contract, or statutory liability if they negligently design systems that produce these discriminatory outcomes. Texas offers a clear warning of the legal consequences and reputational risk that loom large over the deployment of biased AI in the public or private sphere.

In contexts where decisions affect employment, education, lending, or freedom, bias in AI systems can reinforce historical discrimination and undermine public trust.[117] Legal and ethical frameworks must evolve to address these harms by regulating the

113. Texas Responsible Artificial Intelligence Governance Act (TRAIGA), H.B. 149, 89th Leg., Reg. Sess. (Tex. 2025) (to be codified at Tex. Gov't Code § 2050.001) (barring use of AI that infringes on constitutional rights or unlawfully discriminates and authorizing enforcement by Texas Attorney General with civil penalties).

114. *Id.* § 552.105 (authorizing civil penalties of $10,000-$200,000 per violation and $2,000-$40,000 per day for ongoing violations).

115. *Id.* § 552.105 (authorizing civil penalties of $10,000-$200,000 per violation and $2,000-$40,000 per day for ongoing violations). *See also id.* § 2050.106 (providing a 60-day cure period and requiring disclosures for developers and government entities).

116. *See* Houston Fed'n of Teachers v. Houston Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1177 (S.D. Tex. 2017) (holding that use of a proprietary AI to evaluate and terminate teachers may violate due process when its logic is secret and outcomes cannot be challenged).

117. *See generally* EEOC v. iTutorGroup, Inc., JVR No. 2310200016 (E.D.N.Y. Sep. 8, 2023) (settling claims that AI tool unlawfully rejected older applicants in violation of ADEA). *See also* Court Docket, Mobley v. Workday, Inc., No. 3:23-cv-00770 (N.D. Cal. Feb. 21, 2023) (alleging discrimination in AI-based applicant screening software based on race, age, and disability).

fairness and accuracy of AI systems and requiring explainability, accountability, and meaningful avenues of redress.

C. Increasing Business Performance by Eliminating Bias

Although AI can reproduce and amplify historical patterns of discrimination, it can also identify and reduce those same biases.[118] When developed and governed ethically, AI can promote fairness while driving improved business outcomes by expanding markets, enhancing customer trust, and increasing profitability.[119] Given the risks of poorly developed and governed AI systems and the benefits of ethically aligned and transparent design, eliminating bias is not merely a compliance objective or ethical aspiration but a sound business strategy.[120]

Recent advancements in fairness-aware AI have produced practical tools to identify, measure, and mitigate bias throughout the AI lifecycle.[121] Fairness algorithms, for example, are designed to adjust model weights, rebalance training datasets, or modify outputs to meet fairness criteria, such as equalized odds, which ensures similar error rates across protected groups, or demographic parity, which aims to equalize outcomes regardless of sensitive attributes.[122] Pre-processing techniques provide another avenue for mitigation by reweighing data to correct historical imbalances or remove features correlated with protected characteristics to prevent indirect discrimination.[123]

In addition to algorithmic adjustments, organizational structures that incorporate human-in-the-loop mechanisms remain critical. [124] These systems combine machine efficiency with human oversight by allowing individuals to review, audit, or

118. *How AI Can End Bias*, SAP (July 24, 2024), https://www.sap.com/blogs/how-ai-can-end-bias [https://perma.cc/3W4T-N55Z] [hereinafter SAP].

119. *Id.*

120. Claire Duffy, *Lawsuit Claims Discrimination by Workday's Hiring Tech Prevented People Over 40 from Getting Hired,* CNN (May 22, 2025), https://www.cnn.com/2025/05/22/tech/workday-ai-hiring-discrimination-lawsuit (on file with the *Stetson Business Law Review*) (reporting on allegations that Workday's AI-driven hiring tools disproportionately rejected older applicants, illustrating how weak AI governance can expose companies to legal and reputational risk).

121. SAP, *supra* note 118.

122. Ferrara, *supra* note 81, at 4–8.

123. *Id.*

124. INT'L ORG. OF SEC. COMM'NS, *supra* note 15, at 50–52 (discussing how human-in-the-loop mechanisms are essential for maintaining control, trust, and accountability in AI systems used in financial markets).

override AI-generated recommendations. Human involvement at key decision points helps preserve accountability and allows for case-by-case intervention when biased outputs are detected.[125] Moreover, continuous monitoring is essential, as bias can re-emerge over time due to model drift or shifts in underlying data distributions.[126] Ongoing validation and real-world testing ensure fairness commitments are sustained beyond initial deployment.[127]

Legal professionals advising businesses on the creation and use of AI tools should encourage the use of bias impact assessments modeled after privacy impact assessments required under regulatory regimes such as the General Data Protection Regulation (GDPR)[128] and the California Consumer Privacy Act (CCPA), as amended by the California Privacy Rights Act (CPRA).[129] These assessments should document an organization's fairness objectives, evaluate potential disparate impacts, and assess trade-offs between model accuracy and fair outcomes. To be effective, these assessments must be included in a broader AI governance framework that includes periodic audits, internal accountability mechanisms, and stakeholder transparency.

By adopting these tools and practices, businesses can develop AI systems that minimize reputational harm, reduce legal exposure, and tap into previously underserved markets. In the lending context, for example, fairness-oriented models have demonstrated the ability to expand credit access while improving risk management outcomes.[130]

---

125. *Id.*

126. Mirko Bagaric et al., *The Solution to the Pervasive Bias and Discrimination in the Criminal Justice: Transparent Artificial Intelligence,* 59 AM. CRIM. L. REV. 95, 144–46 (2022) (emphasizing that continuous evaluation is necessary to detect and correct bias in deployed AI systems).

127. *Id.*

128. *See generally* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), art. 35, 2016 O.J. (L 119) 1, 53–54 [hereinafter *GDPR*] (stating the procedure for conducting an impact assessment).

129. CAL. CIV. CODE § 1798.185(a)(15) (West 2023) (directing California Privacy Protection Agency to issue regulations requiring risk assessments and cybersecurity audits for businesses whose processing presents significant risks to privacy or security).

130. *See generally* DAVID SCHARFSTEIN & RYAN GILLAND, ZEST AI: MACHINE LEARNING AND CREDIT ACCESS (HARV. BUS. SCH. CASE NO. 9-224-033, rev. June 3, 2024) (examining Zest AI's use of machine learning models to expand credit access while managing regulatory and fairness concerns in lending).

Case Study 3: Fairness in Credit Underwriting – Zest AI

Zest AI provides a compelling example of how prioritizing fairness in AI systems can simultaneously serve ethical and business goals. Founded initially as ZestCash in 2009, the company rebranded as Zest AI in 2019 and shifted from direct lending to licensing machine learning-based credit models to financial institutions.[131] Zest's models are designed to expand credit access for historically underserved populations without sacrificing predictive accuracy.[132] By partnering with credit unions – institutions that prioritize the value they provide to their customers – Zest aligned its fairness-first approach with the mission of its clients, and successfully broadened access to credit.[133]

Unlike traditional credit scoring methods that rely heavily on FICO scores and sparse historical data, Zest AI models are trained on hundreds of features derived from consumer credit bureau data and are customizable to reflect institutional preferences.[134] The company incorporates fairness directly into the modeling process through adversarial debiasing, which systematically reduces the influence of variables that correlate with protected class status.[135] These models undergo disparate impact testing and can be tuned by lenders to select their preferred balance between accuracy and inclusiveness.[136]

Importantly, this ethical AI design yields measurable business benefits. For example, VyStar Credit Union used Zest's platform to automate 75% of lending decisions, increased its portfolio by 22%, and issued $40 million more in new credit annually without increasing its risk exposure.[137] Zest also supports regulatory compliance by automatically generating model risk documentation and using explainability tools like SHAP[138] to make credit

131. *Id.*
132. *Id.*
133. *Id.*
134. *Id.*
135. *Id.* at 7–8 (describing the use of adversarial debiasing to reduce discrimination while preserving model performance).
136. *Id.* at 8–9 (explaining how clients can tune fairness to align with business or regulatory goals).
137. *Id.* at 9 (reporting business outcomes from VyStar Credit Union's implementation).
138. SHAP (SHapley Additive exPlanations) is a model-agnostic method for explaining the predictions of machine learning models. It attributes a model's output to individual inputs, using principles derived from cooperative game theory. SHAP allows users to

decisions auditable. These capabilities have positioned Zest AI as a provider of high-performing models and a trusted partner in navigating regulatory and reputational risk.[139]

Zest's case illustrates that responsible AI governance need not come at the expense of innovation or profitability. On the contrary, when fairness considerations are embedded early in system design and supported by rigorous testing, documentation, and explainability tools, they can become a source of competitive advantage. Ethical AI practices can unlock access to underserved markets, reduce regulatory and litigation risk exposure, and enhance stakeholder trust.[140] For lenders and other high-risk sectors, the path to unbiased outcomes may also be the most strategic path to growth and sustainability.

Eliminating bias is increasingly a business imperative, not merely a compliance or reputation concern. As AI's capabilities expand, so too do the consequences of its misuse.[141] Organizations that invest in fairness-aware systems and proactive governance are better positioned to avoid harm, build consumer confidence, and comply with evolving regulatory standards. Lawyers advising businesses in this space must understand the legal and ethical implications of AI bias and how mitigating that bias can serve their clients' commercial objectives.

IV. REGULATORY FRAMEWORKS, AI GOVERNANCE, AND THE ROLE OF LEGAL PROFESSIONALS

As concerns about algorithmic bias, accountability, and AI-induced harm grow, governments around the world are taking steps to regulate AI more assertively.[142] In both the United States

generate individualized, quantitative explanations of why a particular prediction was made, thereby increasing transparency and aiding in compliance with legal requirements for explainability of automated decision-making. *See* Scott M. Lundberg & Su-In Lee, *A Unified Approach to Interpreting Model Predictions,* 31 ADVANCES IN NEURAL INFO. PROCESSING SYS., 2017, at 4765, 4769–74.

139. SCHARFSTEIN & GILLAND, *supra* note 130, at 5–6 (noting integration of transparency tools and automatic compliance reporting).

140. *See* Andrea Bucher, Comment, *Navigating the Power of Artificial Intelligence in the Legal Field*, 62 HOU. L. REV. 819, 821–22 (2025).

141. Yoshua Bengio et al., *Managing Extreme AI Risks Amid Rapid Progress*, 384 SCI. 842, 842 (2024).

142. Alex Engler, *The AI Regulatory Toolbox: How Governments can Discover Algorithmic Harms*, BROOKINGS (Oct. 9, 2023), https://www.brookings.edu/articles/the-ai-regulatory-toolbox-how-governments-can-discover-algorithmic-harms/ [https://perma.cc/JBZ5-ZHGF].

and abroad, a common theme has emerged calling for AI systems that are transparent, fair, and subject to human oversight.[143] While approaches differ across jurisdictions, legal professionals are increasingly expected to translate these regulations into actionable strategies for their clients.

A. AI Regulation in the United States

In the United States, the regulatory landscape remains fragmented, reflecting a broader pattern of decentralized governance seen in other areas of emerging technology regulation.[144] Much like data breach notification and consumer privacy laws before it, AI governance appears to be headed down a path where dozens of states enact laws in the absence of comprehensive federal legislation.[145]

States and municipalities have entered the regulatory void, enacting sector-specific laws focused on hiring, surveillance, and consumer protection. For example, Illinois passed the Artificial Intelligence Video Interview Act in 2019, which requires employers to provide notice and obtain consent before using AI to evaluate job interviews.[146] New York City's Local Law 144 mandates independent bias audits and transparency for automated hiring tools.[147] Other jurisdictions have recently passed laws that limit the use of facial recognition (e.g., Maryland)[148] or grant job candidates the right to opt out of AI screening (e.g., Colorado).[149]

While no comprehensive AI law exists at the federal level, the Federal Trade Commission (FTC) has asserted that its authority under § 5 of the FTC Act extends to AI systems.[150] In a 2024

143. *Id.*

144. Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy,* 114 COLUM. L. REV. 583, 587–89 (2014) (explaining how U.S. privacy law is fragmented across various sectors, lacks a comprehensive framework, and results in a patchwork of protections depending on industry and jurisdiction).

145. Anjana Susarla, *How States Are Placing Guardrails Around AI in the Absence of Strong Federal Regulation*, THE CONVERSATION (Aug. 6, 2025), https://theconversation.com/how-states-are-placing-guardrails-around-ai-in-the-absence-of-strong-federal-regulation-260683 [https://perma.cc/3Y3R-4QFN].

146. H.B. 2557, 101st Gen. Assemb., Pub. Act 101-0260 (Ill. 2019) (codified at 820 ILL. COMP. STAT. 42/1 et seq.).

147. N.Y.C. Local Law No. 144 (2021) (codified at N.Y.C. ADMIN. CODE § 20-870 et seq.).

148. MD. CODE ANN., Lab. & Emp. § 3-717 (2020).

149. COLO. REV. STAT. §§ 6-1-1701-1707 (2024) (employment AI opt-out).

150. Anthony E. DiResta & Zachary Sherman, *The FTC Is Regulating AI: A Comprehensive Analysis* (July 25, 2023), https://www.hklaw.com/en/insights/publications/2023/07/the-ftc-is-regulating-ai-a-comprehensive-analysis [https://perma.cc/EA7B-RDFN].

enforcement update, the FTC warned that it may treat biased or deceptive AI tools as unfair or deceptive trade practices, subject to investigation and penalties.[151] However, the absence of federal preemption or harmonization has created a patchwork of legal obligations that pose significant compliance challenges for national and multinational businesses.[152]

Whether these early state and municipal initiatives will serve as the foundation for a national AI governance model or be superseded by future federal legislation remains to be seen. In the meantime, legal professionals must guide clients through this evolving and uneven regulatory terrain.

B. International AI Regulation

Outside the U.S., regulators have moved toward more coordinated and enforceable frameworks. The European Union's AI Act, finalized in 2024, imposes tiered requirements based on system risk, with strict obligations for "high-risk" AI used in employment, credit scoring, and public services.[153] These obligations include risk assessments, transparency disclosures, human oversight, and ongoing monitoring.[154] Importantly, the EU AI Act's extraterritorial scope means that U.S. companies marketing AI products in the EU must comply with its provisions or face significant penalties.[155] Complementing the AI Act, the EU's General Data Protection Regulation (GDPR) continues to restrict automated decision-making involving personal data and enshrines the right to explanation and redress.[156]

---

151. Fed. Trade Comm'n, *FTC Announces Crackdown on Deceptive AI Claims and Schemes,* Press Release (Sep. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes [https://perma.cc/PMQ7-D4LG].

152. *Id.*

153. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (*AI Act*), 2024 O.J. (L 206) 1, 16.

154. *Id.* at 19–22.

155. Ana Hadnes Bruder & Arsen Kourinian, *The Impact of the EU AI Act On AI Reseller Deals,* MAYOR BROWN (Nov. 14, 2024), https://www.mayerbrown.com/en/insights/publications/2024/11/the-impact-of-the-eu-ai-act-on-ai-reseller-deals [https://perma.cc/TAB6-J3NC]; Nils Rauer, *The EU AI Act: what US businesses need to know,* PINSENT MASONS (May 17, 2024), https://www.pinsentmasons.com/out-law/analysis/the-eu-ai-act-what-us-businesses-need-to-know [https://perma.cc/YC48-G5YX].

156. GDPR, *supra* note 128, at 14.

Other jurisdictions are also advancing AI governance models. The United Kingdom and Canada emphasize algorithmic accountability and bias mitigation through proposed legislation and guidance for specific sectors.[157] Singapore has taken a leadership role by developing the AI Verify framework, a voluntary set of testing protocols and governance checklists that allow businesses to demonstrate compliance with ethical standards.[158] The UAE, India, and G7 nations have each issued principles or strategies emphasizing fairness, safety, and cross-border cooperation.[159]

Together, these developments reflect a growing consensus around core governance values of transparency, explainability, accountability, and respect for human rights. Yet for multinational companies, the divergence in national laws may present a serious compliance risk. An AI system trained in one country may be deployed in another and used by a company headquartered in a third, raising questions of jurisdiction, enforcement, and conflicting obligations. This global complexity underscores the need for internal governance structures capable of meeting the highest applicable standard, a "most stringent law wins" approach to compliance.

C. AI Governance Structures: Embedding Ethical Guardrails

The National Institute of Standards and Technology (NIST) has made significant contributions to the development of

157. *A Pro-Innovation Approach to AI Regulation: Policy Paper*, U.K. DEP'T FOR SCI., INNOVATION & TECH. (Aug. 3, 2023), https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [https://perma.cc/9HEL-VMCX]; *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. & ECON. DEV. CAN., https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document [https://perma.cc/7Y49-9HB3] (last visited Sep. 21, 2025).

158. *AI Verify Framework,* SING. INFOCOMM MEDIA DEV. AUTH. (July 26, 2024), https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2022/sg-launches-worlds-first-ai-testing-framework-and-toolkit-to-promote-transparency [https://perma.cc/LMW4-6Q25].

159. *National Artificial Intelligence Strategy 2031*, U.A.E. (2019), https://ai.gov.ae/strategy/ (on file with the *Stetson Business Law Review*); *See also Responsible AI: Part 1 – Principles for Responsible AI*, NITI AAYOG (Feb. 2021) https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf [https://perma.cc/S58G-MD2A] (India); *Hiroshima Process International Guiding Principles for Advanced AI Systems*, EUR. COMM'N (Oct. 30, 2023), https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system [https://perma.cc/4FWQ-SZZ9] [hereinafter EUR. COMM'N].

trustworthy AI practices through its AI Risk Management Framework.[160] This voluntary yet influential guidance provides a flexible but rigorous blueprint for assessing and mitigating AI risks.[161] It emphasizes principles such as fairness, transparency, and accountability, and includes practical tools for bias mitigation, documentation, and stakeholder engagement.[162] Although nonbinding, the NIST framework is quickly becoming a de facto standard in the United States and serves as a valuable bridge between the fragmented U.S. regulatory environment and more comprehensive international regimes.[163]

Other organizations have developed complementary frameworks that reinforce these principles on a global scale. The Organisation for Economic Co-operation and Development (OECD) has issued AI principles adopted by over 40 countries, emphasizing human-centered values, robustness, and transparency.[164] The International Organization for Standardization (ISO) has released technical guidance for AI risk management and system lifecycle oversight,[165] while the G7 and G20 have published policy frameworks promoting responsible AI deployment.[166] Legal

160. *See generally Artificial Intelligence Risk Management Framework* (AI RMF 1.0), NAT'L INST. STANDARDS & TECH. (Jan. 2023) [https://doi.org/10.6028/NIST.AI.100-1] [hereinafter NIST AI RMF] (initiative by the U.S. Department of Commerce to implement safe AI standards).

161. *Id.* at 2.

162. *Id.* at 20.

163. *Initial Summary Analysis of Responses to the Request for Information (RFI) Evaluating and Improving Cybersecurity Resources: The Cybersecurity Framework and Cybersecurity Supply Chain Risk Management*, NAT'L INST. STANDARDS & TECH., 1, 8 (June 3, 2022), https://www.nist.gov/system/files/documents/2022/06/03/NIST-Cybersecurity-RFI-Summary-Analysis-Final.pdf [https://perma.cc/NEK6-32Y8] [hereinafter NIST RFI]; *CSF 1.1 International Perspectives*, NAT'L INST. STANDARDS & TECH. (Feb. 26, 2024), https://www.nist.gov/cyberframework/csf-11-international-perspectives#:~:text=%E2%80%9CThe%20adoption%20of%20a%20common,and/or%20conflicting%20expectations.%E2%80%9D [https://perma.cc/S4D5-MMWX] [hereinafter NIST CSF].

164. *See generally Principles on Artificial Intelligence*, OECD, https://oecd.ai/en/ai-principles [https://perma.cc/F57N-PSMK] (promoting international standards on trustworthy AI).

165. *See generally ISO/IEC 23894:2023, Artificial Intelligence – Guidance on Risk Management*, INT'L ORG. FOR STANDARDIZATION (Dec. 2023), https://www.iso.org/standard/77304.html (on file with the *Stetson Business Law Review*) (created in an effort to standardize AI practices); ISO/IEC 42001:2023, *Management System for Artificial Intelligence*, INT'L ORG. FOR STANDARDIZATION (Dec. 2023) https://www.iso.org/standard/81230.html (on file with the *Stetson Business Law Review*) [hereinafter ISO/IEC 42001:2023].

166. *See generally* EUR. COMM'N, *supra* note 159 (adopting principles such as lifecycle risk-based governance and hazard monitoring); *G20 Ministerial Statement on Trade and Digital Economy*, G20 AI PRINCIPLES (2023), https://wp.oecd.ai/app/uploads/2021/06/G20-

professionals advising clients on AI governance can and should draw on these resources to help shape internal compliance protocols and strategic business decisions.

D. The Role of Attorneys

Attorneys are uniquely positioned to shape how AI is governed within organizations, across industries, and under the law. As AI becomes deeply integrated into critical decisions, legal professionals must serve as compliance advisors and strategic partners in designing fair, transparent, and accountable AI systems. Frameworks developed by NIST, the OECD, ISO, and the EU provide a foundation for this work, offering principles and practices that lawyers can help translate into internal policies, procurement standards, and risk management protocols. Whether reviewing contracts with AI vendors, advising corporate boards on algorithmic risk, or guiding clients through impact assessments and audit processes, attorneys must help organizations move beyond minimal compliance and toward ethical AI governance. As regulatory frameworks evolve and public scrutiny intensifies, legal counsel must ensure that AI systems align with the law and values such as justice, accountability, and human dignity.

AI will not wait for legislation to catch up. AI can perpetuate bias, erode civil liberties, and undermine trust when used without sufficient oversight.[167] Therefore, ethical AI governance must be more than just avoiding litigation or reputational damage; it must protect rights, preserve institutional legitimacy, and ensure that technology serves the public good. This moment demands proactive leadership from the legal profession; and attorneys must be ready to meet the challenges presented by AI adoption. By embracing a more forward-looking, multidisciplinary role, legal professionals can help organizations navigate complex regulatory environments, design principled AI systems, and build long-term public trust. The following part offers practical recommendations for how lawyers can take steps toward that goal.

AI-Principles.pdf [https://perma.cc/9R6L-499H] (endorsing OECD AI Principles promoting fairness, accountability, and transparency).

167. Theresa Adie, *Harnessing Technology to Safeguard Human Rights: AI, Big Data, and Accountability*, HUM. RTS. RSCH. CTR. (Apr. 8, 2025), https://www.humanrightsresearch.org/post/harnessing-technology-to-safeguard-human-rights-ai-big-data-and-accountability [https://perma.cc/9L2X-8H2N].

V. RECOMMENDATIONS FOR ETHICAL AI USE IN BUSINESS

Organizations are increasingly relying on AI to drive innovation, improve efficiency, and enhance decision-making processes. This rapid adoption brings heightened risks and compliance challenges that demand proactive solutions. As regulators, clients, and the public place increasing emphasis on transparency, fairness, and accountability in automated systems, businesses must not only avoid legal pitfalls but also build and maintain trust in their use of AI technologies. The following recommendations are designed to equip legal professionals and business leaders with practical frameworks and strategies for ensuring the ethical deployment of AI in the business environment.

A. Proactive Measures for Bias Mitigation

Preventing AI bias is not a one-time event. It requires ongoing monitoring, governance, and intervention throughout the AI lifecycle. Attorneys should advocate for early-stage involvement and long-term accountability. The following recommendations outline practical methods for addressing AI bias issues.

1. Require Algorithmic Audits

Businesses deploying AI systems that impact people, such as hiring, lending, or fraud detection, should perform algorithmic audits before deployment, periodically during use, and after major system updates.[168] These audits can assess disparate impact on protected classes, fairness across demographic groups, data sampling and representativeness, and transparency of model logic and outcomes.[169] Legal departments should ensure that audit results are documented and, where necessary, remediated. In

168. *See* N.Y.C. ADMIN. CODE §§ 20-870–871 (2021) (requiring independent bias audits for automated employment decision tools and transparency measures for candidates).

169. Jeffery Recker, *What is an Algorithmic Bias Audit?,* MEDIUM (Feb. 7, 2023), https://medium.com/@jeffery-recker/what-is-an-algorithmic-bias-audit-ea71252b0ec3 [https://perma.cc/W2RB-NM5W]; Adriano Koshiyama et al., *Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms,* THE ROYAL SOC'Y PUBL'G, May 15, 2024.

regulated industries, these audits may also serve as evidence of due diligence.

2. Embed Transparency Requirements

Transparency should be a default requirement for AI tools, especially when decisions affect individuals' rights and opportunities. Legal counsel should work with business and technical teams to implement explanation mechanisms, utilize disclosure notices that inform users when AI is involved, and develop data lineage documentation that traces how inputs lead to outputs. This is especially important in jurisdictions with emerging laws on automated decision-making, such as the GDPR in the EU, the Colorado Privacy Act, or the California CCPA/CPRA.[170]

3. Use Bias Impact Assessments Early in Development

Like privacy impact assessments, bias impact assessments help businesses evaluate the fairness and social consequences of AI systems early in the design process.[171] These assessments may include a review of historical discrimination risks, consideration of vulnerable populations, and mapping potential legal or reputation harm.[172] Attorneys can help tailor these assessments to align with civil rights laws and evolving regulatory frameworks.

4. Limit Proxy Variables

Even when protected attributes like race or gender are excluded from models, other variables, such as zip code, education level, and even how a person uses language, can act as proxies.[173]

---

170. *See* GDPR, *supra* note 128, art. 22, at 46 (providing data subjects the right not to be subject to automated decision-making without meaningful explanation or recourse); COLO. REV. STAT. §§ 6-1-1301-1314 (2024); CAL. CIV. CODE §§ 1798.100–1798.199.100 (West 2023).

171. *See Advancing Accountability in AI: Governing and Managing Risk Throughout the Lifecycle for Trustworthy AI,* OECD DIGIT. ECON. PAPERS NO. 349 (Feb. 2023), [https://doi.org/10.1787/2448f04b-en] [hereinafter OECD].

172. Jacob Metcalf et al., *Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts,* 739, 742 (2021), https://dl.acm.org/doi/pdf/10.1145/3442188.3445935 (on file with the *Stetson Business Law Review*).

173. Ferrara, *supra* note 81, at 4–6 (discussing the role of proxy variables in perpetuating bias and identifying strategies such as preprocessing and adversarial debiasing to mitigate their impact).

Legal review should include proxy discrimination analysis to identify and mitigate these effects before deployment.

B. Legal Guidelines and Policies: Building Ethical Infrastructure

Legal departments can help organizations create a formal AI policy framework that promotes innovation while managing risk. This framework should be integrated into broader corporate governance and compliance structures. The following recommendations should be considered when establishing a formal AI policy framework.

1. Create Internal AI Use Policies

Internal AI use policies should define permitted and prohibited uses of AI; roles and responsibilities for AI oversight; standards for documentation, explainability, and auditability; and the escalation and incident reporting process. Policies should distinguish between high-risk and low-risk use cases, triggering more rigorous oversight for applications that impact people's legal rights, employment, health, or finances.[174]

2. Integrate AI into Compliance Programs

AI systems should be treated like other regulated processes and be subject to ongoing compliance monitoring. Legal teams should incorporate AI into enterprise risk assessments, align AI practices with internal controls (e.g., ISO framework), and include training on the proper use of AI in employee training programs.[175]

3. Develop a Legal Review Process for AI Use Cases

Just as businesses often have a legal review step for contracts or advertising, they should establish a pre-launch review process for AI deployments. Legal professionals can evaluate consent and data protection requirements, discrimination risks, IP ownership

174. Bradford Kelley et al., *Considerations for Artificial Intelligence Policies in the Workplace,* LITTLER (Mar. 10, 2025), https://www.littler.com/news-analysis/asap/considerations-artificial-intelligence-policies-workplace [https://perma.cc/L8ST-DTFU].

175. *See generally* NIST AI RMF, *supra* note 160 (providing a framework in which employees could be trained).

(especially for AI-generated content), and contractual terms with vendors and third-party providers. This process institutionalizes legal foresight, reducing the likelihood of compliance surprises after deployment.[176]

4. Align with External Standards and Certifications

Legal departments should advocate for the voluntary adoption of external standards, such as the NIST AI Risk Management Framework, ISO/IEC 42001 AI Management System, IEEE's Ethically Aligned Design, or Singapore's AI Verify Framework. In the absence of binding regulation, voluntary alignment with recognized frameworks demonstrates accountability to regulators, consumers, and investors.[177]

C. Cross-industry Collaboration

Ethical AI governance cannot be achieved by lawyers alone; it requires substantial and sustained collaboration between legal, technical, operational, and regulatory stakeholders. Attorneys are uniquely positioned to serve as translators and integrators for these constituencies.[178] The following recommendations allow attorneys to foster such collaboration.

1. Establish Cross-Functional AI Governance Boards

Internal cross-functional AI governance boards should include representatives from legal, compliance, data science and engineering, information security, product management, human resources, and risk and audit committees. Their responsibilities may include reviewing AI use cases, setting ethical standards,

176. *See* OECD, *supra* note 171, at 26.

177. *See* NIST AI RMF, *supra* note 160, at 1; ISO/IEC 42001:2023*, supra* note 165, at § 7.3 Awareness; *Ethically Aligned Design*, INST. OF ELEC. & ELEC. ENG'R, https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf [https://perma.cc/2DS3-GZ9D] (last visited Sep. 21, 2025); *AI Verify Testing Framework*, AI VERIFY FRAMEWORK, https://aiverifyfoundation.sg/what-is-ai-verify/ [https://perma.cc/HM4S-P5CM] (last visited Sep. 21, 2025).

178. *AI Governance: Why In-House Lawyers Need to Lead the Charge,* PLUME, https://www.plume.law/blog/why-in-house-lawyers-need-to-lead-the-charge-on-ai-governance [https://perma.cc/NE2G-EUEA] (last visited Sep. 21, 2025).

monitoring implementation, and coordinating responses to incidents or regulatory inquiries.[179]

2. Encourage External Stakeholder Engagement

Attorneys should support their clients in engaging with regulatory agencies to anticipate and inform AI rulemaking; industry trade groups to share best practices; civil society organizations to understand public concerns and expectations; and academic researchers to incorporate cutting-edge fairness and explainability techniques. Such collaboration enhances public trust, allowing companies to stay ahead of compliance and innovation curves.[180]

3. Promote Cross-Training and Share Vocabulary

Many of the misunderstandings around AI ethics arise from siloed knowledge within an organization.[181] Legal teams can initiate internal cross-training sessions where engineers learn legal basics concerning disparate impact and consent, lawyers gain basic data literacy, and business teams explore real-world ethical dilemmas and case studies. This shared understanding enables better communication, faster risk identification, and more aligned decision-making.[182]

Ethical leadership will become a defining feature of long-term business success as AI matures. Companies that treat AI governance as an ongoing, collaborative, and legally grounded process will be in a better position to innovate responsibly, comply with evolving regulations, and maintain the trust of customers and the public. For attorneys, this presents an opportunity to become strategic partners in shaping how businesses build, deploy, and oversee AI systems, ensuring they are legally compliant and ethical.

---

179. Dan Clarke, *Start Smart: Build an AI Governance Committee and Framework That Scales,* TRUYO (June 26, 2025), https://truyo.com/start-smart-build-an-ai-governance-committee-and-framework-that-scales/ [https://perma.cc/2AC7-KCGJ].

180. OECD, *supra* note 171, at 40-41 (emphasizing the importance of engagement with regulators, industry groups, civil society, and academia to promote transparency, fairness, and innovation in AI governance).

181. James Steinhoff, *AI ethics as subordinated innovation network*, 39 AI & SOC'Y, 1995, 1997–98 (2023).

182. NIST AI RMF, *supra* note 160, at 19.

VI. CONCLUSION

As AI becomes deeply embedded in government, business, and legal practice, attorneys have a responsibility to guide their clients on the development and use of AI in ways that align with human values, ethical norms, and the rule of law. Lawyers must uphold traditional professional duties such as competence, confidentiality, and candor, while extending those duties into emerging AI systems that impact people's lives, liberties, and livelihoods.

This article explored the ethical obligations attorneys face in using and advising on AI, the risks of AI bias, and the role legal professionals play in shaping responsible AI governance. While regulations are still catching up, the tools of the legal profession can be immediately deployed to protect against harm and promote fairness.

To support practitioners in this effort, a practical Ethical AI Governance Checklist is included in the Appendix. In a time of rapid technological transformation, it is incumbent on lawyers to lead in building AI systems that are both innovative and ethical.

APPENDIX

Ethical AI Governance Checklist[183]

1. Before Deployment

 Conduct an AI Use Case Risk Assessment

 Confirm alignment with the client's obligations under privacy laws (e.g., GDPR, CCPA/CPRA)

 Review the AI system for potential bias and disparate impact

 Perform or require algorithmic audits

Check vendor contracts for AI-related indemnities and explainability clauses

183. This checklist was developed by the author based on principles drawn from the NIST AI Risk Management Framework, A.B.A. Resolution 604, and international AI governance recommendations. It is intended as a practical tool for attorneys advising businesses on ethical AI deployment.

2. During Development or Implementation

 Recommend Bias Impact Assessments (BIA)

 Confirm data minimization and appropriate consent

 Assess model transparency: Is it explainable? Can decisions be challenged?

 Document human-in-the-loop processes

 Recommend internal AI governance boards

3. Post-Deployment Oversight

 Monitor for model drift or renewed bias

 Reassess legal risks if use expands or data inputs change

 Provide ongoing training to staff and legal teams about AI capabilities and limitations

 Update compliance policies and incident response plans to account for AI errors