PAPER TITLE: A Legal Perspective on Artificial Intelligence Governance

Joel Wuesthoff (Author/Presenter)
Protiviti
888 7th Avenue
New York, NY 10106
Joel.wuesthoff@protiviti.com

Spencer Judd (Author/Presenter)
Protiviti
101 N Wacker Drive
Chicago, IL 60606
Spencer.judd@protiviti.com

SESSION TITLE: *How to Build Your Firm's AI Governance Framework and IT Architecture*

Presented by:
Joel Wuesthoff, *Protiviti*, New York, NY
Spencer Judd, *Protiviti*, Chicago, IL

Joel Wuesthoff, a former practicing attorney with professional certifications in information security and privacy, serves as a Managing Director and AI Governance Leader for Protiviti Legal Consulting. With over 20 years of experience, he assists global corporations and outside counsel with complex litigation and government investigations involving governance mandates, enterprise-wide compliance obligations, and emerging AI governance requirements.He advises clients on operationalizing AI-related laws and regulations (including the EU AI Act, U.S. regulatory guidance, the NIST AI RMF, and OCC SR 11-7), as well as data protection laws such as GDPR, CCPA, and other emerging automated decision-making frameworks. As a nationally recognized thought leader on transformative legal and organizational issues—such as workplace upskilling and federated AI governance, Joel helps organizations navigate increasing regulatory, privacy, and security mandates. As both a CISSP and former attorney, Joel is uniquely positioned to advise on the legal, operational, and technical implications of emergent AI capabilities and risks, including issues at the intersection of privacy and security regulations.

Spencer Judd is an Associate Director within Protiviti's Technology Audit and Risk practice with over 9 years of experience across a variety of IT audit fields. His experience includes the planning, execution, and reporting of various IT audits and assessments, as well as managing IT Risk within large organizations. He obtained his undergraduate degree from the Mendoza College of Business at the University of Notre Dame. Spencer is a member of the emerging technology capability community within Protiviti's Technology Audit and Risk practice,

supports Protiviti's AI governance and risk management initiatives within internal audit, and is a frequent speaker on AI governance and risk management.

A Legal Perspective on Artificial Intelligence Governance

Joel Wuesthoff| Managing Director – Legal Risk & Compliance
Nicholas You | Director – Legal Consulting

Introduction: AI Governance in a New Legal Reality

When this paper was first published several years ago, AI governance was largely a matter of principle and preparation – a forward-looking exercise in building the right policies before the regulatory and litigation landscape caught up with the technology. That moment of anticipation is over. Today, defense attorneys and their clients are operating in an environment where AI-related lawsuits have moved past motions to dismiss and into discovery, where courts are ordering production of proprietary training datasets and algorithmic parameters, where hundreds of lawyers have been sanctioned for submitting AI-generated fabrications to courts, and where a patchwork of state laws is imposing concrete compliance obligations with real penalties. The foundational governance principles outlined in this paper – transparency, data quality, bias mitigation, and regulatory awareness – have proven prescient. But they must now be understood through the lens of active litigation, evolving case law, and the practical realities of defending clients (and law firms themselves) in an era where AI governance failures carry measurable legal consequences.

The Current Litigation Landscape: What Defense Attorneys Need to Know

Algorithmic Discrimination Has Entered the Courtroom

The most consequential development for defense attorneys is the emergence of class-action litigation targeting AI-driven decision-making tools. In Mobley v. Workday, Inc. (N.D. Cal., Case No. 23-cv-00770-RFL), the court allowed discrimination claims against an AI vendor to proceed on an "agent" theory of liability, holding that the provider of AI hiring tools could be directly liable under Title VII, the ADEA, and the ADA for the discriminatory outcomes produced by its algorithms – even though it was not the employer making final decisions. In May 2025, the court certified a nationwide collective action under the ADEA on behalf of applicants over 40, and by August 2025 had ordered Workday to produce its full customer list, exposing employers across the country to potential discovery obligations regarding their own use of AI screening tools. This is no longer a hypothetical risk. Plaintiffs' counsel are now armed with academic research – including a University of Washington

study showing AI resume screening tools preferred white-associated names in 85% of cases – and are using it to establish systemic patterns of algorithmic bias. Related cases in housing (SafeRent Solutions, settled for $2 million), media (Sirius XM), and other employment contexts confirm that this is an expanding front, not an isolated event. Defense attorneys advising clients who deploy AI tools in hiring, lending, insurance, or housing should be preparing now for interrogatories and document requests targeting how those tools were tested for bias, what human oversight was implemented, and what the organization knew about potential discriminatory impacts.

AI Hallucinations and the Expanding Duty of Verification

Since the Mata v. Avianca scandal in 2023, the number of documented cases involving AI-generated fabrications in legal filings has grown to over 486 worldwide, with 324 in U.S. courts, according to the hallucination database maintained by researcher Damien Charlotin at HEC Paris. The sanctions landscape has escalated sharply. In Gauthier v. Goodyear Tire & Rubber Co. (E.D. Tex., 2024), an attorney was sanctioned for using Claude AI to draft a brief with fabricated citations. In the ByoPlanet litigation (S.D. Fla., 2025), an attorney who submitted hallucinated citations across eight matters – and then used fabricated quotations in response to a show-cause order – saw four federal cases dismissed. Attorneys for Mike Lindell were fined $3,000 each in July 2025 for a filing with over two dozen AI-generated errors. And Morgan & Morgan, the nation's largest personal injury firm, was sanctioned after its own in-house AI platform generated fake case citations. Perhaps most significant for defense practitioners is Noland v. Land of the Free, L.P. (Cal. App., 2025), where the court sanctioned the filing attorney but declined to award fees to opposing counsel who failed to detect the hallucinated citations – suggesting an emerging duty for all attorneys to identify AI fabrications in opposing filings. Over 200 federal judges have issued standing orders requiring AI disclosure. Courts now distinguish between "inadvertent reliance" and "intentional deception," but both result in sanctions. The message for law firms is clear: governance frameworks must include mandatory verification protocols for any AI-assisted work product, and those protocols must be documented to demonstrate defensible practice.

Discovery, Deepfakes, and the Coming Evidentiary Battles

Two additional fronts demand attention. First, courts are now ordering production of AI training data and algorithmic details in litigation. In Tremblay v. OpenAI (N.D. Cal., January 2025), a federal judge ordered production of a complete training dataset used to train GPT-4. In NYT v. OpenAI, source code discovery was permitted only in a secure room disconnected from the Internet. In the OpenAI copyright litigation (S.D.N.Y., May 2025), the court ordered preservation and segregation of all output log data for potential disclosure. OpenAI's CEO has acknowledged that ChatGPT conversations are not privileged communications. Defense attorneys should advise clients that AI interactions – prompts, outputs, and chat logs – may be discoverable, and that inputting privileged information into non-enterprise AI tools may result in privilege waiver. Second, deepfake evidence is creating a dual crisis: fabricated evidence that appears authentic, and authentic evidence that parties claim is fabricated (the so-called "liar's dividend"). The Advisory Committee on the Federal Rules of Evidence released proposed Rule 707 for public comment in August 2025, which would govern machine-generated evidence using expert witness standards for reliability. Louisiana became the first state to establish a deepfake evidence framework, and several cases – including Huang v. Tesla and USA v. Khalilian – are testing authentication standards for AI-generated or AI-challenged evidence. Defense counsel should build deepfake awareness into early case assessment, tailor interrogatories to discover AI-generated materials, and budget for digital forensic expertise.

Foundational AI Governance Principles

The following sections present the foundational governance framework originally published in this paper. These principles – transparency, data quality, bias mitigation, and regulatory awareness – remain the essential building blocks for any organization's AI governance posture, and have been validated by the litigation and regulatory developments described above.

Prioritizing transparency to ensure accountability

Without transparency around how AI systems work and make decisions, it will be impossible to properly govern these technologies or ensure they align with ethical and social mores and values. When AI systems make consequential decisions, the reasoning behind those decisions must be explainable and auditable. Otherwise, it becomes

impossible to hold the right stakeholders accountable if harm occurs, which can have potentially devastating consequences for an organization's reputation. Comprehensive transparency requirements can trace accountability across the full AI lifecycle – from developers building the models, to companies deploying AI to regulators providing oversight.

Transparency also supports effective oversight by an organization and its stakeholders. Auditing AI systems allows an organization to assess factors like data quality, algorithmic bias and model robustness. Access to key technical details through transparency requirements enables proactive governance to identify and mitigate risks early in the AI deployment process. Without such transparency, governance is reduced to reactive crisis management when problems inevitably occur.

Transparency builds public trust and confidence in AI by dispelling notions of "black box" systems. Individuals impacted by AI systems have a right to understand why outcomes were rendered. Understanding the strengths and limitations of AI builds confidence that its risks are being properly managed. Opacity breeds mistrust; transparency enables people to see AI is being developed ethically and deployed safely. Responsible AI developers should consider engineering explainability directly into their models.

Ensuring the quality and privacy of data

Data quality is a keystone for AI governance. For AI to produce insightful and reliable outcomes, the data input into this emerging technology needs to be accurate and relevant. No amount of algorithmic finesse can overcome low-quality training data. Data governance is thus an urgent priority from both legal and ethical considerations. Once a data inventory has been built, categorizing data will allow you to empower informed decision-making. Clean and accurately labeled data empowers AI models to recognize patterns, anticipate trends and offer insights that fuel strategic decisions. In this context, data quality translates directly into business success.

There are complex legal questions related to data privacy when AI systems utilize massive datasets, including personal information, to train their algorithms. What types of data require consent to use? How can personal data be protected from exploitation? How is privacy regulated when data crosses borders? These issues are already creating discussions among regulators and policymakers. In the European Union, the AI Act has now entered its phased implementation period, with obligations for general-purpose AI

models taking effect as of August 2025, and comprehensive high-risk system requirements scheduled for August 2026.

To embark on a successful journey of AI integration within any organization, IT departments and chief compliance officers should begin by compiling an exhaustive list of all products, features and processes that leverage AI. This comprehensive inventory forms the foundation upon which legal and privacy risk management strategies will be built. Leveraging existing data maps or inventories and identifying personal data that AI systems will process is critical to ensuring that data privacy is considered. By uncovering data sources, patterns and connections, organizations can ascertain how AI systems will interact with different types of data.

Identifying and mitigating bias

Another central concern is bias. Historical training data may reflect existing prejudices in society, encoding discriminatory associations directly into machine learning models. Human developers can also inadvertently introduce their own biased judgments into algorithm design choices and data labeling. Unfortunately, bias can be difficult to recognize during development cycles. And once deployed at scale, biased AI tends to amplify rather than mitigate historical inequities. This raises both ethical and legal questions about discrimination and the violation of human rights. From a legal perspective, biased AI may lead to claims of disparate impact under equal protection laws. Impacted groups could argue discriminatory treatment even if bias was unintentional. The onus falls on organizations to be proactive about bias testing before launching AI systems.

To detect and resolve biased AI, transparency is essential but not sufficient. Developers must audit algorithms and training data to uncover hidden biases and understand how they arose. However, auditing alone will not necessarily fix all threats of bias. A thoughtful AI governance framework is needed to provide guidance and oversight for improving the entire development lifecycle.

Navigating the shifting landscape of AI regulations

AI governance is still an emerging policy area, with new laws and regulations being actively drafted and adopted worldwide. By closely following worldwide initiatives and monitoring

regulatory changes, a company can remain compliant. In addition, implementing an AI governance framework with the flexibility to adapt as regulations shift provides a foundation for forward-facing compliance. The core tenets of transparency, accountability and ethics should remain anchors, but methodologies will need to evolve alongside regulations.

Becoming familiar with AI governance and staying up-to-date with legislative updates is essential in the rapidly evolving technological and regulatory landscape. Here are tips to stay informed and knowledgeable:

- Understand the basics of AI, including large language models. There are online courses available that will give you a foundational knowledge of this emerging technology.
- Stay up to date on technological advances in the AI field by subscribing to newsletters, such as the AI News Digest.
- Keep an eye on regulatory developments related to AI governance, such as privacy laws, data protection regulations and AI ethics guidelines set by governments and industry bodies.

AI governance is a monumental challenge, with many open questions and high stakes. As AI becomes more advanced and integrated into our lives, implementing a governance framework with proactive legal perspectives will be crucial in minimizing risk to an organization.

Looking Ahead: Building Defensible AI Governance for Law Firms and Their Clients

The developments described in this paper converge on a single practical imperative: governance is no longer optional, and it is no longer merely aspirational – it is the primary mechanism for legal defensibility. In July 2024, the American Bar Association issued Formal Opinion 512, its first comprehensive ethics guidance on generative AI, establishing that lawyers using AI must fulfill duties of competence, confidentiality, communication with clients, candor toward the tribunal, supervisory responsibility, and reasonable fee practices. State bars in California, Florida, New York, New Jersey, and Pennsylvania have followed with their own guidance. Meanwhile, state legislatures have moved aggressively to regulate AI in employment, housing, insurance, and consumer transactions: in 2025 alone, 38 states adopted approximately 100 AI-related measures, with major laws taking effect in California (October 2025 and January 2026), Illinois (January 2026), Texas (January 2026),

and Colorado (June 2026). These laws impose concrete obligations – bias audits, impact assessments, consumer notice, data retention, and human oversight requirements – that directly parallel the governance principles this paper has advocated from the outset.

For law firms, building a defensible AI governance framework means addressing two audiences simultaneously. For clients, it means being prepared to advise on the discovery demands that now accompany AI litigation – interrogatories targeting algorithm testing, document requests for training data and bias audit results, and preservation obligations for AI chat logs and output data. It means helping clients build the documentation, testing protocols, and oversight structures that will withstand scrutiny in the courtroom. For the firms themselves, it means implementing the internal policies, verification workflows, and training programs that ABA Opinion 512 and the sanctions case law now demand. The firms that have already established AI governance boards, mapped their AI usage (including shadow AI), implemented dual-review protocols for AI-generated work product, and documented their compliance efforts are not only avoiding sanctions – they are winning client confidence in an environment where, as one industry survey found, 60% of in-house legal teams do not know whether their outside counsel is using generative AI on their matters.

The critical insight emerging from two years of litigation, regulation, and real-world AI deployment is that successful AI implementation depends on how well an organization has prepared its information ecosystem to support it. Firms that invested in connecting previously siloed systems, standardizing data processes, organizing their information architecture, and training their people before introducing AI tools are the ones seeing meaningful returns – and avoiding the costly failures that make headlines. The foundational governance principles outlined in this paper – transparency, data quality, bias mitigation, and regulatory awareness – are not abstract ideals. They are, as the case law now demonstrates, the building blocks of defensible practice. The organizations that treat AI governance as a strategic investment rather than a compliance checkbox will be the ones best positioned to harness AI's transformative potential while protecting themselves, their clients, and the integrity of the legal system.

Protiviti is not a law firm and is not providing legal advice or analysis.