SCIENCE ZONE INTERNATIONAL CONFERENCE ON LAW, GOVERNANCE AND SOCIAL TRANSFORMATION VOL 2. N.1 2026

CONFIDENTIALITY IN THE AGE OF ARTIFICIAL INTELLIGENCE: RETHINKING LEGAL DUTIES IN AUTOMATED DECISION-MAKING

Anna Ubaydullayeva
Lecturer of Law and Technology Department,
Tashkent State University of Law
a.ubaydullayeva@tsul.uz

ABSTRACT

The rapid integration of artificial intelligence (AI) and automated decision-making (ADM) systems into professional services has created a profound tension between technological efficiency and the traditional legal duty of confidentiality. This article examines the shifting landscape of professional secrecy, specifically focusing on how large language models and autonomous algorithms challenge the established frameworks of attorney-client privilege and data protection. Through an analysis of current regulatory responses, including the European Union's AI Act and recent professional ethics opinions, the study identifies critical vulnerabilities in data retention, model training, and third-party access. The research concludes that traditional "notice and consent" models are insufficient for the opaque nature of AI processing. Instead, a "zero-trust" governance framework is proposed, necessitating a redefinition of legal duties to include algorithmic transparency and proactive data redaction as foundational components of modern professional ethics.

KEYWORDS

Artificial Intelligence, Confidentiality, Automated Decision-Making, Legal Ethics, Data Privacy, Professional Duty

Introduction

The evolution of professional practice has reached a critical juncture where the deployment of artificial intelligence is no longer speculative but an operational necessity. As firms increasingly rely on automated decision-making to handle vast datasets, the foundational principle of confidentiality faces unprecedented technical challenges. Traditionally, confidentiality was maintained through physical security and controlled digital access; however, the probabilistic nature of modern AI introduces risks that transcend these legacy boundaries.

The core problem lies in the architectural design of generative AI and ADM systems, which often require the ingestion of sensitive data to function or improve. This inherent "data hunger" creates a direct conflict with the legal duty to protect client information from unauthorized disclosure. Legal professionals, in particular, find themselves navigating a paradox where the tools used to enhance their competence may simultaneously jeopardize the very privilege that defines their role.

This article explores the systemic risks associated with AI-driven workflows, focusing on the potential for inadvertent disclosure during model training and the erosion of "privileged" status when data is shared with third-party AI providers. By evaluating the legal duties of professionals in 2026, this study seeks to provide a roadmap for rethinking confidentiality in an era where the "watcher" is often an invisible algorithm.

Literature Review

Recent scholarship has begun to map the complex intersection of AI and privacy, noting that while many problems are variations of longstanding issues, AI "remixes" them in unique ways (Solove, 2025). Previous frameworks focused on static data storage, but current research highlights the dynamic risk of "memorization" in large language models (LLMs), where systems may inadvertently reproduce sensitive training data in subsequent outputs (LeanLaw, 2025). This phenomenon effectively turns a private input into a potential public disclosure.

Furthermore, the legal status of AI providers as "third parties" has become a central point of contention in privilege doctrine. Historically, sharing

information with outside entities not essential to representation waived attorney-client privilege. Some scholars argue for an "agency doctrine" where AI is treated like a paralegal or translator, yet this requires a level of control and confidentiality that many consumer-facing AI platforms explicitly disclaim in their terms of service (Walter, 2025).

Regulatory responses have also evolved. The European Union's AI Act (2024/2026) has introduced a risk-based approach, categorizing certain applications, such as credit scoring or judicial assistance, as "high-risk" (Gartner, 2026). These systems are now subject to strict transparency and audit requirements. Despite these advancements, a gap remains between high-level regulation and the granular ethical duties required of individual practitioners.

Methods

This study employed a qualitative analysis of current legal frameworks, professional ethics opinions issued between 2024 and 2026, and recent case law regarding data breaches in AI systems. The primary source material included the American Bar Association (ABA) 2024 guidance, the EU AI Act, and reports from the 2026 India-AI Impact Summit.

A comparative approach was used to evaluate how different jurisdictions—specifically the United States, the European Union, and India—are balancing the "regulation-innovation" dichotomy. The analysis focused on four key vectors of risk: unauthorized data retention for model training, third-party cloud vulnerabilities, the "black box" nature of algorithmic decision-making, and the failure of traditional consent mechanisms in the face of complex AI architectures.

Results

The findings indicate that the standard of "reasonable steps" to prevent disclosure is rapidly shifting toward a requirement for technical literacy. In 2025, it was observed that while 31% of legal professionals utilized generative AI, only 21% of firms had formal policies in place to govern its use (LeanLaw, 2025). This gap represents a significant liability, as most consumer-facing AI

tools do not provide the confidentiality guarantees required for privileged communications.

Data suggests that breaches involving AI platform credentials or "prompt leakage" take an average of 328 days to identify and contain, leading to prolonged exposure of sensitive client strategies (IBM, 2025). Moreover, the transition to "AI-generated news" and automated summaries has created a "nutrition label" movement, where tech companies are pressured to disclose the provenance of the data used in their outputs to prevent the dilution of professional standards (IPPR, 2026).

In the judicial context, the duty to state reasons is being challenged by ADM systems. Research indicates that judges using AI for drafting judgments must now disclose such reliance to satisfy the right to a fair trial, as established under Article 50 of the AI Act (Juliussen, 2025). This highlights a shift from confidentiality being a passive state to an active duty of transparency regarding the "human-in-the-loop" involvement.

Discussion

The implications of these findings suggest that the traditional understanding of confidentiality as "non-disclosure" is no longer sufficient. Professionals must now adopt a "zero-trust" posture, where no AI system is presumed to be secure by default. This involves a move away from public, consumer-grade tools toward specialized, "sandboxed" AI environments where data is not used for model training and remains within the firm's sovereign control.

A rethinking of the "notice and consent" model is also necessary. Because the internal logic of AI is often opaque, clients cannot truly give "informed" consent to the use of their data in such systems. Instead, the burden of ensuring confidentiality must fall on the deployer to implement "privacy by design," utilizing tools like automated redaction of personally identifiable information (PII) before any data is processed by an external model (Bergman, 2026).

Furthermore, the legal profession must address the "memorization risk" through a lifecycle obligation. Approval of an AI tool cannot be a one-time event; it requires ongoing monitoring to ensure that as the model evolves or "believes its

own fiction" (Gartner, 2026), it does not develop new vulnerabilities that could lead to the exposure of confidential patterns or strategies.

Ultimately, the goal is to integrate AI in a way that preserves the "ancient" ethical obligation of trust while harnessing modern efficiency. This requires a harmonized effort between regulators, tech providers, and professionals to create "human-centered" AI governance that prioritizes fundamental rights over speed to market.

Conclusion

Confidentiality in the age of artificial intelligence requires a fundamental shift in how legal duties are conceptualized and executed. The move from manual to automated decision-making does not absolve the professional of their duty; rather, it intensifies the need for technical oversight and proactive risk management. By embracing zero-trust governance and algorithmic transparency, professionals can safeguard the foundation of the client relationship.

Future research should focus on the development of "privacy-preserving" AI technologies, such as federated learning and differential privacy, which may offer a technical solution to the confidentiality paradox. For now, the best practice remains clear: professionals must redact first and leverage AI second, ensuring that the "compass" of ethics continues to guide the "ship" of innovation.

Would you like me to expand on the specific technical measures, such as differential privacy, that could be used to enhance these confidentiality frameworks?

References

Bergman, R. (2026). *Why legal professionals must use redaction tools before using AI*. Mediate. https://mediate.com/protecting-clients-preserving-privilege-and-avoiding-liability-why-legal-professionals-must-use-redaction-tools-before-using-ai/

Gartner. (2026). *What should we do when AI starts believing its own fiction?* LiveMint. https://www.livemint.com/newsletters/tech-talk/what-should-we-do-when-ai-starts-believing-its-own-fiction-11769740508304.html

Juliussen, K. (2025). *Rethinking the judicial duty to state reasons in the age of automation?* Cambridge Forum on AI Law and Governance. https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/rethinking-the-judicial-duty-to-state-reasons-in-the-age-of-automation/0984E85BC2519D5E5E448FAFCCBD98F6

LeanLaw. (2025). *AI Privacy Risks: Protecting client data in 2025*. https://www.leanlaw.co/blog/what-are-the-data-privacy-implications-of-using-ai-tools-with-confidential-client-information/

Solove, D. J. (2025). *Artificial intelligence and privacy*. Florida Law Review. https://www.floridalawreview.com/article/129976-artificial-intelligence-and-privacy.pdf

Walter, A. N. (2025). *Artificial intelligence and attorney-client privilege*. Walter Counsel. https://waltercounsel.com/artificial-intelligence-and-attorney-client-privilege/