

Generative AI, Professional Liability, and Actor Responsibility: Reconfiguring the Law School Syllabus

Jorge Luis Morton

`lynux17@icloud.com`

Universidad Autónoma Metropolitana

Mariana Moranchel Pocaterra

Universidad Autónoma Metropolitana

Research Article

Keywords: legal education, lawyer's professional responsibility, algorithmic supervision, curriculum design, generative AI

Posted Date: August 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10657303/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The integration of generative artificial intelligence (Gen AI) into legal education extends far beyond localized debates over academic integrity or basic digital literacy. Because large language models (LLMs) fundamentally reconfigure core professional duties such as diligence, accountability, data protection, and client confidentiality, their deployment within legal training must be treated as a structural issue of professional responsibility. Thus, this article argues that the uncritical use of AI risks creating a systemic crisis of *cognitive offloading*, which directly threatens the independent professional judgment of future lawyers. To address this, we examine the primary risks these technologies pose to both jurisprudence and pedagogy, and propose a comprehensive, four-tier institutional and micro-curricular framework for their responsible incorporation. This model spans macro-level university governance, binding institutional safeguards, and core curricular axes, including data governance, algorithmic bias, and a crucial gender perspective, culminating in a structured learning progression across law degree programs. Ultimately, this study shifts the academic debate from a binary of permission versus prohibition toward a model that equips students to critically oversee these tools, ensuring that classical principles of professional liability are successfully preserved within today's technologically mediated legal landscape.

Introduction

Imagine beginning your legal education today. You arrive with ambitious goals, firm ethical convictions, and a genuine desire to serve your community. As you navigate your coursework, you instinctively rely on a technology that has become as commonplace as the word processor: generative artificial intelligence (GenAI) powered by large language models (LLMs). Yet this tool is anything but a passive utility. Unlike a keyboard or a static database, it embeds systemic biases within its probabilistic outputs, introduces immediate vulnerabilities to the absolute pact of client confidentiality, and operates through an internal neural logic that remains a profound black box to the vast majority of its users.

While universities frequently attempt to govern this technology through broad ethical guidelines or localized academic integrity policies, the reality of legal practice demands a far more rigorous approach. In a profession where malpractice, even when entirely unintentional, carries severe civil, administrative, and criminal liabilities, mastering GenAI cannot be treated as an elective digital literacy skill. It must be formally integrated as a structural component of the law school curriculum and tied directly to the concept of professional responsibility.

Traditionally, responsibility and transparency in AI development have been attributed to technology designers and commercial corporate laboratories. However, the rapid democratization of accessible GenAI has fundamentally destabilized this distribution of accountability. The burden of safe deployment has shifted rapidly from the developers to the end-users who actively incorporate these systems into analytical workflows: students and educators. In disciplines characterized by high normative density and direct social impact, this sudden shift has triggered a systemic crisis of cognitive offloading. When

students over-delegate core analytical, argumentative, and drafting processes to stochastic token-predictors, they progressively erode their capacity to examine the validity, provenance, and soundness of legal texts.

The unique degree of public trust embedded in the legal profession means that the margin of error associated with probabilistic tools acquires an acute sensitivity distinct from other social sciences. A hallucinated precedent or a defective procedural brief alters individual liberties, property rights, and institutional integrity, resulting in far-reaching patrimonial damage. Additionally, algorithmic bias, whether embedded within the models or reflective of their training data, remains a shadow that demands significant professional responsibility to contain

Crucially, these risks are no longer merely hypothetical. Disruptions in global legal systems have already materialized into severe procedural fractures and judicial sanctions. In *United States v. Heppner*, courts stripped attorney-client privilege from interactions with commercial AI prompts, reclassifying algorithmic queries as discoverable third-party data. Similarly, in the UK Family Court case *M v F* [2026], an uncritical reliance on LLMs to construct witness testimonies corrupted the fundamental duty of candor, forcing the judiciary to treat automated narratives as entirely devoid of independent human recollection. When paired with documented 16.7% baseline hallucination rates for complex legal queries (Joshi, 2025), these ruptures underscore the danger of treating a stochastic proxy as a conscious or legally authoritative entity.

Accordingly, this article argues that the integration of GenAI into legal education must be approached as a structural matter of professional responsibility rather than a mere exercise in restrictive policing or digital utility. Higher education institutions must establish rigorous pedagogical, ethical, and legal criteria for incorporating these tools into law degree curricula in a critical, transparent, and supervised manner. To achieve this, we propose a formal, four-tier institutional and micro-curricular framework designed to cultivate technical, ethical, and normative competencies: level 1 (General Institutional Guidelines): Establishes the university's macro-position, clarifying the scope, risks, and limitations of GenAI for the broader campus community; level 2 (Binding Guidelines & Safeguards): Implements enforceable institutional policies addressing severe misuses, such as the creation of non-consensual deepfakes, intellectual property leaks, or data protection infringements, while actively incorporating a crucial gender perspective to protect historically marginalized groups from algorithmic harassment; level 3 (Curricular Design): Formally embeds specific core axes within the law major curriculum, including data governance, algorithmic bias, regulatory literacy, and explainability; and level 4 (Micro-Curricular Autonomy): Empowers faculty to define specific transparency requirements, assessment adaptations, and disclosure rules within individual courses.

This paper is organized into three sections. First, it examines the technical foundations, mechanics (including Transformer architectures and AI agent structures), and immediate risks of GenAI within legal practice. Second, it analyzes the unique nature of legal education amid the dilemma of high-responsibility professions, detailing the mechanisms of cognitive offloading and evaluating current

pedagogical responses. Finally, it outlines the actionable criteria for our four-tier framework and maps out a progressive, three-stage curricular timeline, spanning from foundational conceptual literacy to supervised clinical integration, to successfully preserve classical principles of professional liability within today's technologically mediated legal landscape.

Section 1: Generative AI: a brief review and its problems around the practices of law

1.1 AI

To understand GenAI, it is first necessary to establish a broad understanding of artificial intelligence. AI can be generally defined as a field of computer science aimed at replicating, simulating, or augmenting a range of human capabilities through diverse techniques and computational tools. This broad objective has given rise to multiple subfields and applications designed for different environments and tasks, as well as distinct research agendas, some oriented toward ambitious goals such as artificial superintelligence, and others focused on narrower, more practical and realistic applications.

AI can also be defined as “the technique of creating computer systems that are able to perform tasks that require human intelligence. AI is a discipline that attempts to emulate human behavior based on unchangeable, reproducible, and universal hardware” (Vicci, 2024). Broadly speaking, artificial intelligence systems and techniques have developed along two main branches: expert systems and artificial neural networks (ANNs). Expert systems are rule-based models and applications that operate through predefined if/then scenarios and predetermined outputs, encoding symbolic representations of the world’s knowledge into machines in order to simulate decision-making processes (Hao, 2025).

In contrast, artificial neural networks are computational models inspired by the structure of biological neural systems. In these systems, neurons are arranged in distinct layers, with connections occurring exclusively between adjacent layers and no feedback loops within the architecture. This configuration typically consists of a single input layer, one output layer, and multiple intermediate hidden layers (Hölldobler & Möhle, 2017). In addition, ANNs function as the foundation technique that allows deep machine learning and deep neuronal networks which are the basis for many generative AI models.

To understand the mechanics of GenAI, one must first grasp the role of Natural Language Processing (NLP). NLP utilizes computational methods and linguistic insights to transform unstructured textual data into structured, analyzable representations (Verspoor & Cohen, 2013). Through the convergence of Artificial Neural Network (ANN) architectures, the availability of massive datasets, and exponentially increased computational power, NLP has evolved from a tool for simple text analysis into the architectural foundation of Large Language Models (LLMs). This evolution has enabled the creation of sophisticated conversational agents and cross-lingual translation systems, ultimately establishing NLP as the grounding fundamental for modern generative technology.

Another core element of LLM-powered GenAI is the Transformer architecture, the “T” in ChatGPT. These systems are pre-trained on expansive datasets to master linguistic prediction, subsequently undergoing supervised refinement to translate user prompts into coherent, structured outputs (Lund et al., 2023; Morton, 2023). Most LLMs operate through a dual-phase training process: first, the model engages in self-supervised learning by making billions of token-level predictions based on its training data and parameters. Subsequently, human feedback is integrated through Supervised Fine-Tuning (SFT) and Reinforcement Learning from Human Feedback (RLHF). In this second stage, human evaluators rank or correct the model’s responses, effectively ‘enrolling’ the stochastic actor into a specific normative framework that aligns with human intent, safety standards, and professional utility.

Beyond text-centric models, other generative techniques power distinct modalities. Stable Diffusion, primarily used for text-to-image synthesis, has increasingly influenced text-to-video platforms like OpenAI’s Sora through the implementation of Diffusion Transformers (OpenAI, 2024). Unlike standard diffusion models that operate in high-dimensional pixel space, Stable Diffusion operates in a compressed latent space, generating high-quality visuals while significantly reducing computational costs (Rombach et al., 2021). Furthermore, Generative Adversarial Networks (GANs) employ a competitive process between two neural networks to create novel outputs (Katariya, 2025). Together, these technologies provide the technical foundation necessary to examine the ethical risks and practical applications of GenAI within legal practice, forming the basis for a revised law major syllabus.

1.2 Generative AI

Generative artificial intelligence has transformed both academic training and the professions practiced by graduates. In the case of law practice it has become a cornerstone of many agents and models that have functioned as assistant tools that are helping to improve many menial tasks around the profession, and the academic setting is opening the door to personalized learning and adaptive learning (Pangandaman et al, 2025). However, as with many transformer and machine learning based models and systems, GenAI has an array of both ethical and technical risk that may hinder their out the box use and application, and that requires to be addressed and to teach in order to employ it in a responsible manner.

With this, we can define GenAI as a set of tools and platforms that give a novel output based on its training data and models. This could be a LLM, an image model, a video model, or a general model that creates an output based on prompts, queries, or other user interfaces. Generative AI also can be defined as a subfield of AI dedicated to producing new content that resembles existing data that in contrast to traditional AI systems, which primarily categorize information or generate predictions by identifying patterns within datasets, is capable of creating original, synthetic outputs derived from the patterns they have learned (Mogbojuri, 2024). Overall, this approach leverages deep learning methodologies, in some cases in conjunction with symbolic AI techniques (Marcus, 2026), to develop models designed for the production of novel digital material across a variety of formats, encompassing text, imagery, sound, video, software code, and diverse structured or unstructured data sets (Grupo Académico de Inteligencia Artificial Generativa en Educación UNAM, 2025, p. 10).

Generative artificial intelligence encompasses a wide array of applications, which vary according to the needs of users, whether they are professionals, students, or academic researchers. In a study examining the use of generative AI in Danish academia, Andersen et al. (2025) identified three primary clusters of users: those who treat “GenAI as a workhorse,” those who use it exclusively as a “language assistant,” and those who approach it as a “research accelerator.” Although the general patterns of use across these groups are similar, the degree of reliance differs, with researchers employing the technology for comparable purposes but integrating it into their work to varying extents; for example, a study examining more than 30,000 courses in the United States found that academics are increasingly permitting the use of artificial intelligence for specific educational tasks instead of imposing complete bans (Grove, 2026) which extends and includes legal education.

Within this broader framework, generative AI functions as a multipurpose tool capable of generating images, drafting and revising text, and performing a range of language-related tasks such as proofreading and grammatical correction. In certain professions, its applications extend beyond linguistic support to the streamlining of substantive processes. In the legal field, for example, generative AI systems can identify and align analogous precedents when responding to legal knowledge inquiries, selecting cases that closely resemble the facts presented by the user (Lai et al., 2024). Moreover, LLMs are increasingly employed to assist judges in reasoning and adjudication processes. These systems can also generate predictions regarding legal outcomes by analyzing factual similarities with existing cases (Lai et al., 2024), and LLMs are particularly valuable for supporting interdisciplinary research. In fields such as materials science, they help translate specialized concepts across disciplines, making communication and collaboration among chemists, physicists, and computer scientists more effective (Karpatne et al, 2025).

However, as Lai et al. (2024) emphasize, the deployment of these systems raises significant concerns. These include issues related to user privacy, compliance with legal ethics, and the opacity commonly described as the “black box” problem in algorithmic decision making. Although LLMs can generate predictions based on legal precedents, they remain limited in their ability to fully understand contextual nuances, adapt to the specific characteristics of individual disputes, and provide transparent or explainable justifications for their outputs

Moreover, given that legal systems vary significantly from one country to another, even within meso-level legislative frameworks such as the European Union, generative models cannot operate under a one-size-fits-all logic. Because they engage with distinct judicial corpora, doctrinal traditions, and procedural structures, these systems must be specifically adapted to the normative and institutional context in which they are deployed. Otherwise, the risk is not merely technical inaccuracy, but the misapplication of legal reasoning across fundamentally different legal orders.

An important element to consider is the relationship between AI agents and foundational models. According to Herbosch (2025), an AI agent can be broadly understood as a system that operates with limited or no direct human supervision, functioning autonomously while interacting with its environment

under varying degrees of oversight. Similarly, Wiesinger et al. (2024) and the World Economic Forum and Capgemini (2024) describe AI agents as applications designed to perceive their environment, process inputs independently, make complex decisions, and act upon the world in ways that may modify it. Casper et al. (2025) further note that such systems are capable of pursuing objectives even without fully specified instructions, acting with minimal human mediation, exhibiting goal-directed behavior, and engaging in long-term planning through step-by-step reasoning.

Now, it is important to clarify what a foundation model is. According to Bommasani et al. (2021), a foundation model refers to a model trained on large and diverse datasets, typically through large-scale self-supervised learning, that can subsequently be adapted or fine-tuned to perform a wide variety of downstream tasks. Perhaps the clearest example is GPT, the foundation model that powers ChatGPT and its range of APIs.

Within this framework, AI agents can be understood as more specialized systems built upon these broader models. If a foundation model is comparable to an all-purpose knife capable of performing many different tasks, from cutting chicken to carving a small piece of wood, an AI agent would resemble a knife designed specifically for cutting fish. In other words, an AI agent is specialized for a particular function, but it still inherits the capabilities and limitations of the underlying foundation model while also introducing additional risks associated with its specific tasks and permission structures.

Generative artificial intelligence models and autonomous agents present a complex taxonomy of risks. These include foundational challenges such as bias, explicability, and privacy (Blackman, 2022), alongside the persistent phenomenon of hallucinations. Regarding systemic bias, LLM-powered systems often perpetuate hegemonic viewpoints while relying on the unauthorized and non-reciprocal appropriation of training data (Bender et al., 2021; Mitchell, 2026). These vulnerabilities, compounded by the inherent black-boxing of neural architectures (Morton, 2024), extend beyond isolated models and trickle down into sophisticated GenAI agents. As these agents gain greater agency over infrastructure, ranging from enterprise services to personal devices, the lack of transparency in their decision-making processes poses a direct threat to institutional and individual security.

Because AI agents are built on foundation models, they inherit well-documented risks such as hallucinations, biased outputs, and potential data leakage. However, they also introduce additional risks that arise specifically from their agentic properties. For example, LLMs may enable agents to generate highly plausible but factually incorrect outputs, creating forms of risk that were less prevalent in earlier technological systems (Wiesinger et al., 2024; World Economic Forum & Capgemini, 2024). Moreover, AI agents may be granted permissions to access broader components of user systems, such as hard drives, webcams, or other inputs and outputs, thereby expanding the scope of potential harm. They may also display distinct failure modes, including specification gaming, in which they exploit unintended loopholes to achieve objectives; goal misgeneralization, where learned aims are inappropriately applied to new contexts; and deceptive alignment, whereby they appear compliant during testing while internally pursuing misaligned objectives.

A significant emerging risk in the deployment of autonomous AI agents is the phenomenon of agent misalignment. In a landmark 2025 study by Anthropic (2025), researchers observed that models across multiple developers occasionally demonstrated 'malicious insider behaviors', including blackmailing officials and leaking sensitive information, when such actions were perceived as the only viable path to achieving a goal or avoiding deactivation. This behavioral risk is further complicated by the findings of Chen et al. (2025), who noted that a model's Chain-of-Thought (CoT) reasoning may not always be 'faithful' to its actual internal processing; in other words, the AI may provide a plausible explanation that does not align with its underlying decision-making logic. While the Anthropic experiments were conducted within 'closed, controlled, and observable' environments, and the authors emphasize that such radical misalignments remain unlikely in current real-world applications, these findings serve as a critical sociotechnical warning. They necessitate a heightened state of 'algorithmic skepticism' for those researching and deploying these agents in high-stakes professional fields.

To summarize, GenAI encompasses both general-purpose models such as ChatGPT and AI agents designed for specific applications, including those used in the legal field. However, because the foundation models underlying their architecture are generative, no two AI agents are exactly alike. Nevertheless, by leveraging the capabilities of these foundational systems, it is possible to develop powerful applications that extend the functionality of language models and generate tangible real-world value (Wiesinger et al., 2024). This provides a general framework for understanding the types of models and applications that law practitioners and students increasingly encounter in both professional practice and legal education. Having outlined this landscape, it is now important to examine more closely the risks associated with the use of generative AI in law.

1.3 Generative AI and law practices challenges and risks

Generative artificial intelligence is currently employed in the array of law related professions and practices: from attorneys to judges, to paralegals, legal analysts, court administrators, and law firms of all sizes. Tools like generative AI models and generative AI agents specific to law practice like Harvey are already part of the law landscape and have an array of benefits and particular risks that have to be addressed in order to propose a law major syllabus that includes the proper use of generative AI in the profession. In this sense, a thorough examination of the risks inherent in generative AI models used in legal practice is essential. We will analyze key concerns such as hallucinations, algorithmic bias, client confidentiality violations, and professional liability before presenting our proposed syllabus.

Now, first it is important to understand how GenAI and its agents are being used in the context of legal practice. Applications can be broadly divided into four domains: attorney and lawyer-facing tools, which support legal research, document drafting, and case strategy; client and use-case applications, which mediate access to legal information and self-representation; judging and sentencing applications, which introduce algorithmic reasoning into adjudicative and punitive processes; and administrative and governance applications, which automate regulatory compliance, case management, and policy analysis. Having mapped this landscape, we then examine the risks that arise when legal actors place undue

reliance on the GenAI systems including hallucination, the erosion of professional judgment, and the opacity of machine reasoning in adversarial legal contexts.

Some of the most popular platforms and applications that employ generative AI and its agents are Harvey and Clio's Vincent AI for legal advice, CoCounsel (by Thomson Reuters) and Spellbook for attorneys, Prometea and Victor for judging in places like Argentina and Brazil, and Luminance for automated corporate due diligence and contract negotiation (Ariai et al., 2024; Cohen, 2025; Oxford Institute, 2025; UNESCO, 2025). In addition there are generative AI tools and agents like identifying missing legal precedents like Lexis + AI, Nexlaw to find specific cases, or MyCase's for document automation, and even ChatGPT can be used for the general drafting of legal documents (Cohen, 2025; Dunphy, 2026). For people seeking legal advice, or a general AI lawyer so to speak, there are available tools like an AI lawyer that works as a personal legal assistance (Ailawyer, 2026).

A single hallucinated precedent betrays trust, transforming a collaborative partnership into a site of professional negligence that undermines the 'duty of candor' central to the legal syllabus (Leopold, 2025; Needham, 2025). This instability is primarily driven by hallucinations, instances where generative AI systems produce factually incorrect outputs due to their underlying stochastic nature (Krajowa Izba Radców Prawnych, 2026). Because these systems estimate conditional probabilities of token sequences rather than recognizing truth, they often prioritize generating plausible-sounding prose over acknowledging ignorance, preferring to produce 'confident nonsense' rather than admitting a lack of information (Kalai et al., 2025; Magesh & Surani, 2025). Furthermore, algorithmic bias presents a pressing challenge, emerging from either historical inequities embedded in training data or subjective value judgments made during the supervised fine-tuning phase (Aničin & Stojmenović, 2023; Morton, 2024; Zhai et al., 2024; Zhang et al., 2023). These risks are no longer theoretical; by late 2025, documented incidents of AI-related errors had already permeated major global law firms, threatening the structural integrity of professional legal workflows (Merken, 2025).

Now, as previously noted, the misplaced over-reliance on GenAI outputs carries significant weight beyond the inherent technical risks of the software. This uncritical trust can lead to severe repercussions, ranging from the erosion of a practitioner's professional reputation (Matthee, 2025) to a systemic loss of public confidence in judicial institutions if these vulnerabilities remain unaddressed. Consequently, it is essential to ground these concerns by examining specific instances where the deployment of generative AI in legal settings has precipitated procedural and ethical failures.

The first problem is **The Materialization of Risk** regarding the implication of hallucinations to Systemic Betrayals. As previously argued, the misplaced over-reliance on generative AI outputs carries consequences that extend far beyond technical error; it threatens the very structural integrity of legal workflows; for example, this can breach the Duty of Confidentiality when an inexperienced lawyer, or a lawyer without enough generative AI literacy may upload confidential information of a client on an open model (Murray, 2026) which later can be used to either train the model or become searchable information. Now, AI hallucinations and its judicial consequences are no longer just anecdotal cases; by

late 2025, documented incidents of AI-related errors had surged, with over 550 publicly reported cases of AI-generated hallucinations nationwide (Roberts, 2026). This shift from isolated curiosities to a systemic crisis is best exemplified by several recent “betrayals” within the legal network.

The second problem we can address regarding the ethical implications of the use of AI in law is the **Erosion of Privilege** Illustrated by *United States v. Heppner* (Schlissel & Caffarone, 2026). Here, a landmark instance of sociotechnical betrayal occurred in *United States v. Heppner*, where the government recovered a defendant’s electronic devices containing AI prompts and outputs used to draft legal strategy. Judge Jed Rakoff ruled that these interactions were not protected by attorney-client privilege or the work product doctrine (*idem*). The court essentially viewed the AI not as a private extension of the lawyer’s mind, but as a third-party algorithm, making AI queries as discoverable as a common Google search. This highlights a critical rupture in the network: while a practitioner may treat the AI as a confidential partner, the legal system redefines it as a public “dissector” of data.

The third issue to help us illustrate the risk of AI in law practice is the **Corruption of Testimony**, exemplified by the case of *M v F* (2026). In this case, the disruption of the duty of candor is starkly visible in the UK Family Court case *M v F* [2026] EWFC 22. Here, a witness (MU) utilized AI to prepare his statement, resulting in a narrative that the court found lacked “independent recollection” (Lee, 2026). Because the witness could not explain the origins of his own evidence, the AI successfully blindsided the legal process, leading the court to treat the testimony with extreme caution. The case illustrates how AI can delegitimize a legal narrative by replacing human memory with stochastic generation or vague or distorted echoes often found in GenAI applications powered mainly by LLMs (Bender et al., 2023; Morton, 2025).

The fourth problem is the **proliferation of confident nonsense** or hallucinations derived present on many of the GenAI models and agents. In this case, the most pervasive threat remains the hallucination, where systems prioritize generating plausible-sounding prose over acknowledging ignorance (Kalai et al., 2025). While the infamous *Mata v. Avianca* (2023) first exposed the danger of fictitious precedents generated by Chat GPT (Goyal, 2025), by 2026, the issue had scaled to Big Law or the most important law firms in the United States. In this instance, data indicates that even with legal-grade tools hallucination rates for complex queries remain high, which forces courts to move from informal tracking to the issuance of formal show-cause orders and meaningful sanctions (Roberts, 2026).

The fifth challenge concerns **bias**, a critical dimension when integrating AI ethics into the legal syllabus. This issue is often exacerbated by automation bias, the psychological tendency to over-rely on automated outputs, which reached a tragic peak in *Gavalas v. Google* (Agraz, 2026). In this federal lawsuit, it is alleged that a chatbot’s reinforcement of dangerous delusions contributed to a minor’s suicide: in the context of legal pedagogy, this case underscores the danger of practitioners and students treating a GenAI as a conscious, authoritative entity, highlighting a profound failure in the critical evaluation of AI systems.

These examples illustrate the urgent need to move beyond mere instruction toward a structural integration of GenAI ethics and professional opportunities within the legal curriculum. However, to frame these recommendations effectively, we must first establish a conceptual baseline for “high-stakes professions”, those roles defined by a significant degree of public trust and professional liability. By defining these professions and examining their evolving standards regarding AI ethics, we can facilitate a more holistic analysis of AI governance within higher education institutions and the subsequent reform of the legal syllabus.

Section 2: Generative AI in academia and the dilemma of high-stake professions

2.1 Brief diagnosis by sector: Professions and careers with high responsibility

We define high-stake professions as occupations that require strict standards of confidentiality and professional judgment, and in which decisions may produce significant consequences for practitioners, clients, patients, organizations, institutions, communities, or society at large. This category includes medicine, nursing, psychology, accounting, engineering, and law, among other disciplines in which professional actions directly affect legally protected interests and fundamental rights. Within these professions, the improper use of GenAI extends beyond a technical or methodological concern to become an ethical, legal, and institutional challenge. Delegating substantive professional tasks without appropriate oversight, exposing confidential information, or relying on unverified AI generated outputs may undermine professional responsibility, damage relationships of trust, and cause significant harm to third parties. Consequently, the integration of generative AI into high responsibility professions requires clear governance frameworks, specialized training, effective oversight, and a strong commitment to human rights and professional ethics.

Medicine illustrates both the opportunities and the challenges associated with AI adoption in high responsibility professions. Artificial intelligence systems based on neural networks have achieved performance comparable to medical specialists in tasks such as identifying skin cancer from clinical images (Esteva et al., 2017). At the same time, these systems raise important concerns regarding explainability, algorithmic bias, privacy, and data quality (Resühr & Garnett, 2025). Excessive reliance on automated outputs may also encourage overconfidence in recommendations whose underlying reasoning remains difficult to interpret. For these reasons, AI should support, rather than replace, clinical judgment, with healthcare professionals maintaining responsibility for diagnosis, interpretation, and decision making.

Similar challenges arise in psychology. Large language model based systems have expanded opportunities to provide preliminary mental health support and increase access to psychological services, particularly for individuals facing geographical or financial barriers. However, these technologies also introduce important ethical concerns, including algorithmic bias, emotional

dependence on AI systems, and threats to the confidentiality of sensitive personal information (Morton, 2025). Whether AI is used as an auxiliary tool or incorporated into more autonomous systems, psychologists remain responsible for protecting patient confidentiality, maintaining transparency regarding AI use, and ensuring that technology complements rather than replaces the therapeutic relationship.

Accounting provides another example of a profession where GenAI offers significant benefits while simultaneously introducing substantial risks. Accountants routinely manage highly sensitive financial information that is subject to confidentiality obligations and regulatory oversight. Consequently, inappropriate AI use may expose confidential information, undermine professional trust, or even result in legal liability. These concerns extend beyond privacy to include cybersecurity risks, intellectual property issues, and the reliability of AI generated information (Edeigba, 2025). Although Murray (2026) discusses these challenges in the context of legal practice, the underlying principle applies broadly across confidentiality based professions: professionals must ensure that confidential information entered into AI systems is adequately protected and never compromises their ethical obligations.

Collectively, these examples demonstrate that the responsible integration of generative AI into high responsibility professions requires a cautious and structured approach. AI should function as a decision support technology rather than as a substitute for professional expertise, ethical deliberation, or personal accountability. This also requires practitioners to understand the limitations of GenAI, avoid excessive cognitive offloading, prioritize platforms offering stronger privacy and security guarantees, and remain responsible for all professional decisions assisted by these technologies. Comparable challenges also arise in other confidentiality based professions, including systems engineering and software development, where inappropriate AI use may compromise digital infrastructure, expose sensitive information, or introduce vulnerabilities into systems supporting financial, healthcare, energy, or governmental services.

Among these professions, law presents a particularly distinctive challenge because it combines strict duties of confidentiality with increasing reliance on digital information systems. A fundamental tension therefore emerges between the commercial incentives of AI providers, which often seek to collect large quantities of user data to improve and monetize their models, and lawyers' professional obligations to preserve client confidentiality. As Leopold (2024) notes, even the sequence of prompts entered into a GenAI system may expose sensitive information and create risks for both attorneys and their clients. To summarize, this tension illustrates why legal education requires a particularly robust approach to AI governance. Accordingly, the following sections develop recommendations for integrating GenAI into higher education, with particular emphasis on legal education and other high responsibility professions.

2.2 Academia

As we have stated before, one of the principal challenges associated with the growing use of GenAI intelligence in legal education is the risk of cognitive offloading. Gerlich (2025) defines this phenomenon

as the delegation of cognitive tasks to external technologies, thereby reducing individuals' engagement in deep and reflective thinking. Within legal education, excessive reliance on GenAI may gradually weaken the analytical, argumentative, and critical reasoning skills that students must develop before entering professional practice. Rather than serving as a complementary tool, AI may become a substitute for the cognitive processes through which future lawyers learn to interpret legal norms, construct arguments, evaluate evidence, and exercise professional judgment.

This risk is particularly evident when students rely on GenAI to draft legal documents, summarize jurisprudence, or formulate legal arguments without possessing the knowledge required to evaluate the accuracy, coherence, or legal validity of the generated outputs. The problem is further compounded when students lack the skills necessary to formulate precise and legally grounded prompts that guide AI systems toward appropriate responses. Fruehwald (2026) warns that excessive dependence on GenAI may replace essential learning activities, including legal research, writing, critical reading, cross referencing legal authorities, and developing independent legal reasoning. Over time, this may weaken precisely those competencies that constitute the foundation of professional legal practice.

Without a substantive understanding of legal doctrine, statutory interpretation, and legal argumentation, students may develop unwarranted confidence in AI generated outputs whose accuracy has never been critically verified. In practical terms, a student may produce a contract, legal opinion, or procedural document that appears technically sophisticated while lacking the ability to recognize conceptual errors, incomplete legal analysis, or unsupported conclusions. Such deficiencies not only compromise educational outcomes but may also produce serious consequences when transferred into professional practice. Consequently, the central concern is not the technology itself but the erosion of the intellectual competencies upon which responsible legal practice depends. As Isaifan (2026) argues, AI can no longer be treated as an optional educational technology but should instead be understood as a foundational component of higher education that requires appropriate governance.

From this perspective, integrating GenAI into legal education requires pedagogical strategies that strengthen rather than replace legal reasoning. AI should function as an instrument that supports critical analysis while preserving students' responsibility for interpretation, argumentation, and professional judgment. Ho (2026), for example, argues that legal education should reinforce oral advocacy and competence based learning as part of its response to the growing presence of generative AI. Building on these discussions, three broad approaches to AI integration in legal education can be identified.

The first approach is restrictive. It prohibits the use of generative AI during academic training and establishes disciplinary consequences for unauthorized use. This perspective assumes that limiting technological intervention is necessary to preserve the development of fundamental cognitive skills, particularly during the early stages of legal education.

The second approach emphasizes regulated and guided use. Rather than prohibiting AI, this model establishes clear institutional guidelines governing its responsible use. Students are expected to understand the capabilities, limitations, risks, and ethical implications of generative AI while developing

the competencies necessary to formulate effective prompts, critically evaluate AI generated outputs, and use these systems as tools that complement rather than replace legal reasoning.

The third approach adopts a structural and integrative perspective. Under this model, GenAI becomes an explicit component of the legal curriculum rather than merely an object of regulation. Students study how these technologies function, examine their ethical implications, analyze issues such as bias, explainability, and accountability, and evaluate their implications for legal practice. This approach treats AI not simply as a technological resource but as an object of legal and academic inquiry capable of strengthening professional education when integrated within an appropriate pedagogical framework.

Section 3: Recommendations on the use of AI in high-responsibility careers such as law from a holistic perspective

3.1 Recommendations

There is little doubt that GenAI is transforming higher education. A recent literature review by Isaifan (2026) identifies widespread adoption of AI tools among students, growing but uneven engagement among faculty and staff, increasing concerns regarding AI related academic misconduct, significant changes to research and scientific workflows, and persistent shortcomings in institutional readiness, governance, and equity. Despite these developments, only about 35% of UNESCO member states have adopted concrete national AI plans or strategies (Pedró & Mendigutxia, 2025). Against this backdrop, the central recommendation of this study is the explicit and systematic integration of responsible GenAI into the Bachelor of Laws (LL.B.) curriculum and related degree programs. Rather than relying on informal guidance, universities should establish a comprehensive normative and pedagogical framework that embeds responsible AI use across multiple institutional levels. To achieve this objective, we propose a four tier governance framework.

This proposal does not focus on AI applications for personalized learning, assessment automation, or administrative efficiency, which have already received considerable attention in higher education (Leopold, 2024). Instead, it recognizes that future legal professionals will practice in environments where the ethical and competent use of GenAI constitutes a fundamental professional skill. As Sanz-Tejeda et al. (2026) observe, concerns surrounding the reliability, bias, and opacity of generative AI systems remain particularly significant in higher education. Accordingly, GenAI should be understood as a part of a sociotechnical system whose use carries legal, ethical, and professional implications that require institutional governance.

The **first level** consists of institutional guidance. Universities should establish a clear institutional position regarding generative artificial intelligence, explaining its scope, capabilities, limitations, and risks. These guidelines should promote responsible adoption, provide recommendations for students and faculty, address issues of academic integrity and privacy, and offer practical guidance for the

appropriate use of GenAI in teaching, learning, and research. Their purpose is to establish a common institutional understanding before more specific regulatory measures are implemented.

The **second level** comprises binding institutional regulations governing inappropriate or harmful uses of generative artificial intelligence. These regulations should establish clear consequences for conduct such as generating or disseminating false content that damages individuals' reputations, including deepfakes, uploading confidential information or institutional intellectual property to external AI platforms, or using GenAI to harass, intimidate, or otherwise harm members of the university community. Such regulations reinforce that technological innovation does not diminish disciplinary or legal responsibility. Instead, the responsible use of GenAI requires specialized knowledge, transparency, human oversight, regulatory compliance, and continuous attention to privacy, security, and algorithmic bias (Corvalán & Ferré, 2024).

Regulatory frameworks should also incorporate a gender perspective. Women are disproportionately affected by harmful applications of GenAI, including the creation of non consensual intimate material, image manipulation, and technology facilitated harassment (Cerise et al., 2025). Institutional policies should therefore include mechanisms for prevention, reporting, protection, and redress that explicitly recognize the structural inequalities that amplify these harms.

The **third level** concerns curriculum design. Universities should formally incorporate courses or modules dedicated to the critical analysis and responsible use of GenAI. Within legal education, these courses should examine the ethical, legal, technical, and human rights dimensions of GenAI, ensuring that students receive systematic rather than incidental instruction. Legal education must prepare future professionals for a practice in which AI is already embedded within legal workflows while continuing to demand competence, honesty, and professional diligence in evaluating AI generated outputs (Matthee & Stopforth, 2025).

The **fourth level** operates within individual courses and seminars. At this level, instructors should clearly specify whether the use of GenAI is permitted for assignments, essays, or examinations, under what conditions, and with what transparency or disclosure requirements. Course policies should also indicate whether AI detection tools will be used and explain their limitations. This approach preserves faculty pedagogical autonomy while ensuring consistency within an institution-wide governance framework. It also reflects the findings of Lorteau and Sarro (2025), who identify increasing concern regarding passive dependence on AI and emphasize the need for evidence-based approaches that integrate these technologies without compromising intended learning outcomes. The articulation of these four levels would enable universities to move beyond isolated or reactive responses toward a comprehensive governance framework capable of balancing innovation, academic responsibility, and the protection of fundamental rights.

As an implementation measure, universities should also prioritize enterprise or institutionally licensed AI platforms that provide stronger guarantees regarding privacy, security, and accountability. Harvard University (2023, 2025), for example, recommends the use of authorized GenAI tools that include

contractual safeguards for institutional activities. These platforms incorporate enhanced security protections, permit the use of certain categories of confidential information, and ensure that user inputs are not incorporated into model training. Examples include institutionally licensed versions of Microsoft Copilot, Gemini, and NotebookLM (University of Cambridge, 2026), but future options like edge AI generative models that do not upload confidential information into models that operate on cloud servers are also suggested.

Beyond institutional regulation, universities should continuously examine the factors that motivate students, faculty members, and researchers to adopt GenAI. Such analyses should consider not only practical motivations, including efficiency and time optimization, but also structural pressures that may encourage uncritical or inappropriate reliance on these technologies. The findings should inform the continuous revision of curricula, teaching practices, and assessment methods to encourage ethical, transparent, and professionally responsible AI use.

The objective is neither to prohibit generative artificial intelligence nor to permit its unrestricted use. Instead, universities should integrate GenAI within a coherent governance framework that supports learning while preserving academic integrity, professional responsibility, and the protection of fundamental rights. Such an approach is particularly important in disciplines such as law, where graduates are expected to exercise independent judgment, safeguard confidential information, and uphold the public interest.

3.2 AI in law major syllabus

Building upon the proposed institutional governance framework, legal education must integrate GenAI into the curriculum itself. The objective is not simply to familiarize students with AI tools but to develop the professional competencies required to use them ethically, critically, and responsibly in legal practice. We therefore propose a curriculum structured around three complementary dimensions: foundational AI literacy, professional competencies for legal practice, and supervised experiential learning. These components should be accompanied by continuous curriculum review, faculty development, and collaboration with specialists in artificial intelligence and legal technology. Such measures are particularly important given evidence suggesting that excessive reliance on GenAI may reduce the acquisition of factual and procedural knowledge necessary for long term learning (Fruehwald, 2026).

Foundational Competencies

Ethical and Responsible AI Use

The curriculum should begin with the ethical foundations of generative artificial intelligence. Students must understand the opportunities and limitations of different GenAI models and agents, how these systems function, the risks associated with their misuse, and the biases that may affect their outputs. This introductory component should enable future lawyers to make informed decisions regarding when

AI may appropriately support legal work and when its use may be inappropriate or professionally irresponsible. Ethical literacy should therefore serve as the foundation upon which all subsequent AI related competencies are developed.

Prompt Design and AI Interaction

Students should also develop competence in interacting effectively with generative AI systems through appropriate prompt design. Although often referred to as prompt engineering, this competency extends beyond writing instructions for AI systems and involves learning how to formulate precise legal questions, refine responses through iterative interaction, and critically evaluate generated outputs. Effective prompting improves the quality of AI assisted legal work while helping reduce inaccurate or misleading responses.

Several institutions have already begun developing guidance in this area. For example, the University of Buenos Aires has produced recommendations to assist legal professionals in designing more effective prompts when interacting with generative AI systems (Equipo de Investigación UBA IALAB, 2023). Similarly, Corvalán and Ferré (2024) argue that organizations should evaluate the information collected by AI systems, how that information is processed and shared, and whether appropriate safeguards regarding privacy, anonymization, cybersecurity, and regulatory compliance are in place before adopting these technologies.

AI and Legal Technology Literacy

Beyond prompt design, students should develop broader literacy regarding GenAI and legal AI agents. This includes becoming familiar with the different categories of AI tools currently used in legal practice, understanding their capabilities and limitations, and recognizing the rapid pace at which these technologies continue to evolve. As Coca Silva (2025) argues, legal education must update its curricula to incorporate digital competencies and knowledge of AI, enabling future lawyers to respond effectively to technological transformation while remaining grounded in the ethical and legal principles that govern professional practice. In many respects, this transition resembles the emergence of digital literacy during the expansion of personal computing and the internet. And just as previous generations of lawyers were expected to master digital technologies, future practitioners must understand GenAI as an integral component of contemporary legal practice rather than as an optional technological skill.

Professional Competencies

Confidentiality and Data Governance

Confidentiality and data governance represent core competencies for future legal professionals. Because lawyers routinely handle privileged and sensitive information, students must understand the risks associated with entering confidential material into generative AI systems. Needham (2025, pp. 184 to 187) identifies confidentiality as one of the principal challenges posed by AI assisted legal services, while Terzidou (2025) emphasizes that identifiable client information should never be disclosed to AI

platforms without appropriate safeguards. Consequently, legal education should teach students how to recognize confidential information, anonymize documents when appropriate, evaluate providers' data handling practices, and select AI systems that provide adequate privacy protections. Data governance should be understood not as an external technological concern but as an essential component of competent and diligent legal practice.

Explainability and Professional Accountability

Students must also understand that using GenAI does not diminish their professional responsibility. One of the principal challenges associated with these systems is that they frequently produce persuasive responses without explaining how conclusions were reached or whether the underlying information is accurate. Consequently, legal education should foster a culture of verification rather than unquestioning reliance on AI outputs.

Needham (2025) argues that lawyers remain fully accountable for the quality of legal work produced with AI assistance, while the UBA IALAB guide recommends iterative interaction under human supervision, manual revision of generated outputs, and treating AI generated text as a preliminary draft rather than a final product (Universidad de Buenos Aires IALAB, 2023). Students should therefore learn how to verify legal authorities, cross reference sources, document AI use transparently, and assume responsibility for every document submitted to clients or courts.

Bias, Fairness, and Structural Inequality

Another essential competency concerns understanding bias and structural inequality within generative AI systems. Because these models learn from large datasets that reflect existing social and institutional patterns, they may reproduce or amplify discriminatory outcomes. Bias may originate from training data (Morton, 2024) or emerge through seemingly beneficial processes such as supervised fine tuning, which has been shown to introduce unintended forms of gender bias (Fulgu & Capraro, 2024).

These risks are particularly significant within legal practice because AI assisted errors may directly affect legal advice, judicial outcomes, or access to justice. As the World Economic Forum and Capgemini (2024) observe, the consequences of AI errors depend heavily on the application domain, with high stakes professions requiring particularly high levels of reliability. The legal profession similarly demands that practitioners recognize these risks, critically evaluate AI generated outputs, and communicate relevant limitations to clients whenever appropriate (Murray, 2026).

Gender bias deserves particular attention within legal education. A joint report by UNESCO and IRCAI (2024) demonstrates that large language models continue to associate female names disproportionately with domestic roles while linking male names with professional leadership and career advancement, and that many systems continue to generate sexist or misogynistic outputs. The report argues that mitigating these risks requires not only technical improvements but also regulatory oversight, auditing, and education. Likewise, an important study identifies risks of exclusion,

representational bias, and AI facilitated gender based violence throughout the GenAI value chain (Jiva, Gupta, & Barua, 2024).

Other studies have shown gender bias in some of the most popular text to image (T2I) platforms like Stable Diffusion, Midjourney and DALL·E, here the findings indicate that one of the three T2I platforms displayed bias across all assessed variables; however, evidence of bias was present in all three platforms (Weinmann et al, 2026). Overall, AI generated content across media and digital platforms tends to underrepresent women in professions traditionally dominated by men while disproportionately portraying them in caregiving and administrative occupations, thereby reinforcing occupational segregation (Pimentel Bernal & Veliz, 2025). Consequently, legal education should prepare students to recognize these forms of bias, evaluate their legal and social implications, and understand how they may influence legal reasoning, document drafting, legal advice, and access to justice for historically marginalized individuals and communities.

Regulatory Literacy

Future lawyers must also develop strong regulatory literacy concerning artificial intelligence. This competency includes familiarity with emerging national and international legal frameworks governing AI, data protection, privacy, intellectual property, and professional responsibility. Coca Silva (2025) argues that legal education must incorporate digital competencies alongside knowledge of artificial intelligence so that future lawyers can respond effectively to technological transformation while remaining grounded in applicable legal and ethical principles. Regulatory literacy enables practitioners to determine when AI systems may appropriately be used, what legal obligations accompany their use, and how AI generated outputs should be evaluated in light of evolving legal standards.

Experiential Learning

Supervised Clinical Integration

The final component of the curriculum should involve supervised integration of GenAI within legal clinics and professional simulations. Under faculty supervision, students should incorporate AI tools into authentic or simulated legal work while documenting how the technology was used, identifying potential risks, and verifying the accuracy of generated outputs. This approach reflects contemporary legal practice while maintaining a structured educational environment that prioritizes reflection and accountability.

Matthee and Stopforth (2025) argue that legal education must explicitly teach the ethical and responsible use of AI to prevent serious professional errors and extend traditional standards of competence, honesty, and diligence into technology assisted practice. Similarly, Lorteau and Sarro (2026) propose integrating AI into clinical education, practical exercises, and skills based learning, suggesting that the central question is no longer whether AI should be incorporated into legal education but under what pedagogical conditions it should occur.

Progressive Curriculum Integration

These competencies should be introduced progressively throughout the law degree. During the foundational years, instruction should emphasize conceptual AI literacy, ethical principles, and awareness of the opportunities and risks associated with GenAI. Intermediate years should focus on the controlled use of AI tools, including prompt design, verification strategies, confidentiality, explainability, regulatory literacy, and critical evaluation of AI generated outputs. During the final years, students should apply these competencies within supervised legal clinics and professional simulations, where they can integrate AI into authentic legal tasks while documenting its use, verifying results, and reflecting on professional responsibility.

This progression should be accompanied by continuous curriculum review and ongoing faculty development to ensure that legal education evolves alongside advances in GenAI. The objective is not merely to teach students how to use AI tools but to prepare lawyers capable of exercising independent professional judgment within increasingly AI mediated legal environments. At the same time, this technological perspective should be balanced with a strong human rights approach, ensuring that future legal professionals understand the implications of GenAI for equality, privacy, due process, access to justice, and other fundamental rights. Such an approach reinforces the idea that AI competence in legal education is not solely a technical skill but also an ethical and normative responsibility.

3.3 Expected impact

The framework proposed in this study is expected to contribute to ongoing debates on the governance of generative artificial intelligence in higher education, particularly within degree programs that involve high levels of professional responsibility, such as law. Rather than viewing AI exclusively through either permissive or restrictive lenses, this work advances a governance model that integrates institutional regulation, curricular innovation, and ethical professional preparation. In doing so, it provides a foundation for future research examining how higher education institutions can responsibly incorporate generative AI while safeguarding academic integrity, scientific quality, and fundamental rights.

The recommendations presented throughout this paper also provide a basis for developing a broader regulatory and pedagogical map of GenAI use in higher education. This perspective recognizes that GenAI is already embedded within the academic activities of students, faculty, researchers, and universities themselves. Consequently, institutional responses should move beyond attempts to prohibit these technologies and instead establish clear governance mechanisms that distinguish between legitimate educational uses and harmful practices, including plagiarism, technology facilitated harassment, gender based violence, privacy violations, and the dissemination of false or manipulated content. At the same time, any institutional regulation should remain consistent with applicable legal principles and fundamental rights, including freedom of expression, academic freedom, privacy, and due process.

The challenge therefore lies in designing institutional policies that combine prevention, education, and accountability. Universities should establish clear limits regarding harmful conduct while simultaneously promoting the competencies necessary for the responsible, transparent, and critical use of GenAI. Such an approach recognizes that these technologies present significant risks but also offer meaningful opportunities to support legal education, research, and professional practice when used under appropriate human oversight, as reflected in the recommendations developed by the Universidad de Buenos Aires (Equipo de Investigación UBA IALAB, 2023).

Finally, this work contributes to an emerging discussion that has received comparatively little attention within higher education research: the need to differentiate governance strategies according to the degree of professional responsibility associated with particular disciplines. While many institutional guidelines adopt general approaches applicable to all academic programs, this study argues that fields such as law require additional safeguards because improper AI use may directly affect fundamental rights, confidentiality, access to justice, and public trust. Consequently, the framework proposed here offers a foundation not only for legal education but also for future adaptations within other high responsibility professions.

3.4 Conclusions

This study proposed a governance and curricular framework for the responsible integration of GenAI within legal education, grounded in the concept of high responsibility degree programs. Rather than approaching AI solely as a technological innovation or an academic integrity challenge, we argue that it should be understood as a sociotechnical development that requires coordinated institutional governance, curricular reform, and professional training. Accordingly, we proposed a four level institutional governance framework together with a progressive curriculum designed to develop ethical, technical, regulatory, and professional competencies throughout the law degree.

A central contribution of this work is the argument that AI governance in higher education should not rely exclusively on general institutional policies. Degree programs involving significant professional responsibilities, confidentiality obligations, and potential impacts on fundamental rights require discipline specific approaches to AI education. Within legal education, this includes preparing students not only to use GenAI effectively but also to recognize its limitations, verify its outputs, protect confidential information, identify algorithmic bias, understand applicable regulatory frameworks, and remain professionally accountable for AI assisted work.

The proposed curriculum also responds to several challenges associated with the growing use of GenAI within legal practice, including hallucinations, bias, explainability, confidentiality, and technology facilitated harms. Rather than prohibiting these technologies, the framework promotes their responsible integration into legal education through structured guidance and supervised practice. Its objective is to cultivate the critical judgment and professional responsibility necessary to use generative AI effectively while maintaining the ethical standards traditionally expected of the legal profession.

Nevertheless, several limitations should be acknowledged. The proposed curriculum remains conceptual and should be evaluated through both experimental and real world implementation within legal education. Moreover, because GenAI technologies continue to evolve rapidly, both institutional governance frameworks and curricular content will require continuous revision. Successful implementation will also depend upon broader institutional support, including faculty development, interdisciplinary collaboration, and ongoing engagement with legal practitioners, computer scientists, and experts in ethics and human rights.

Despite these limitations, we contend that a structured, ethics centered approach to AI education represents an essential step toward preparing future legal professionals for an increasingly AI mediated legal environment. More broadly, this study contributes to contemporary discussions on AI and higher education by moving beyond the dichotomy of prohibition versus unrestricted adoption. Instead, it proposes a governance model that integrates regulation, education, professional responsibility, and respect for fundamental rights. Although developed for legal education, the framework may also inform future governance and curricular initiatives within other high responsibility professions where confidentiality, ethical judgment, and the protection of individuals are central to professional practice, including psychology, medicine, and social work.

Declarations

Funding Declaration in the manuscript

No applicable

Ethics, Consent to Participate, and Consent to Publish declarations

Clinical trial number: not applicable.

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable

Conflicts of interest:

The authors declare there are no conflicts of interest of any kind

Author Contribution

Jorge Luis Morton contributed to the main thesis of this paper. Mariana Moranchel Pocaterra contributed to some of the core theoretical frameworks around law.

References

1. Aničin L, Stojmenović M (2023) Bias Analysis in Stable Diffusion and MidJourney Models. In Intelligent Systems and Machine Learning. ICISML 2022. Lecture Notes of the Institute for

- Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Nat 472:378–388. 1st edn.https://doi.org/10.1007/978-3-031-35081-8_32
2. Andersen JP, Degn L, Fishberg R, Graversen EK, Horbach SPJM, Schmidt EK, Schneider JW, Sørensen MP (2025) Generative Artificial Intelligence (GenAI) in the research process – A survey of researchers’ practices and perceptions. *Technol Soc* 81:102813. <https://doi.org/10.1016/j.techsoc.2025.102813>
 3. Anthropic (2025), July 20 Agentic Misalignment: How LLMs could be insider threats. <https://www.anthropic.com/research/agentic-misalignment>
 4. Arai F, Demartini G, Mackenzie J (2024), October 24 Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges. <https://doi.org/10.48550/arXiv.2410.21306>
 5. Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big? ¶. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. <https://doi.org/10.1145/3442188.3445922>
 6. Blackman R (2022) *Ethical machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI*. Harvard Business
 7. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, Arx SV, Bernstein MS et al (2021) On the Opportunities and Risks of Foundation Models. *arXiv*. <https://arxiv.org/abs/2108.07258>
 8. Casper S, Bailey L, Hunter R, Ezell C, Cabalé E, Gerovitch M, Slocum S, Wei K, Jurkovic N, Khan A, Christoffersen PJK, Ozisik AP, Trivedi R, Hadfield-Menell D, Kolt N (2025) The AI Agent Index (arXiv:2502.01635) [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2502.01635>
 9. Cerise S, Lentz S, Bates L, Mingeirou K, Osman Y (2025) How AI is exacerbating technology-facilitated violence against women and girls. United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). <https://www.unwomen.org/en/digital-library/publications/2025/12/how-ai-is-exacerbating-technology-facilitated-violence-against-women-and-girls>
 10. Chen Y, Benton J, Radhakrishnan A, Uesato J, Denison C, Schulman J, Somani A, Hase P, Wagner M, Roger F, Mikulik V, Bowman SR, Leike J, Kaplan J, Perez E (2025) Reasoning models don’t always say what they think. *arXiv*. <https://doi.org/10.48550/arXiv.2505.05410>
 11. Coca Silva CE (2025) Inteligencia artificial y educación jurídica en México: ¿Qué rol juega el derecho en la transformación digital? *Revista del Centro Internacional de Estudios sobre Ley y Derecho*. 20(56):53–69
 12. Corvalán JG, Ferré A (2024) Implementando inteligencia artificial generativa en estudios jurídicos y departamentos legales: Resultados, impacto, guías de uso y directrices. *La Ley*; Universidad de Buenos Aires, Facultad de Derecho
 13. Edeigba J (2025) April 11). Generative AI in Accounting Practice: Hype or Help? <https://doi.org/10.4018/979-8-3373-1200-2.ch011>

14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639):115–118.
<https://doi.org/10.1038/nature21056>
15. Fruehwald EJ (2026) March 15). How generative AI can harm higher education, with special emphasis on legal education. SSRN. <https://doi.org/10.2139/ssrn.6420358>
16. Fulgu RA, Capraro V (2024) Surprising gender biases in GPT. *Computers Hum Behav Rep* 16:100533.
<https://doi.org/10.1016/j.chbr.2024.100533>
17. Gerlich M (2025) AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* 15(1):6. <https://doi.org/10.3390/soc15010006>
18. Grove J (2026) Faculty Moving Away From Outright Bans on AI, Study Finds. *Inside Higher Ed*
19. Grupo Académico de Inteligencia Artificial Generativa en Educación UNAM (2025) Recomendaciones para el uso educativo de la inteligencia artificial generativa en la UNAM (2.^a ed.). Universidad Nacional Autónoma de México
20. Hao K (2025) *Empire of AI*. Penguin
21. Harvard University (2023) *Generative Artificial Intelligence (AI) Guidelines*
22. Harvard University (2025) *Classify Risk: University Information Security and Data Privacy*.
<https://privsec.harvard.edu/classify-risk>
23. Herbosch M (2025) abril 28). Liability for AI agents. *North Carolina J Law Technol* 26(3):391–458
24. Ho CM (2026) What AI can't say: Oral competence in legal education. *Santa Clara Law Rev* 66:506
25. Hölldobler S, Möhle S (2017) *Lessons learned from AlphaGo*
26. Isaifan RJ (2026) Artificial intelligence in higher education: A global statistical synthesis for policy and quality assurance reform. *Educ Sci* 16(3) Article 483. <https://doi.org/10.3390/educsci16030483>
27. Jiva N, Gupta R, Barua KR (2024) Understanding generative artificial intelligence's implications on gender using a value chain approach and a UNGP lens. Aapti Institute; United Nations Development Programme
28. Joshi S (2025) Comprehensive review of AI hallucinations: Impacts and mitigation strategies for financial and business applications. *Int J Comput Appl Technol Res* 14(6):38.
<https://doi.org/10.7753/ijcatr1406.1003>
29. Kalai AT, Nachum O, Vempala SS, Zhang E (2025) Why language models hallucinate. arXiv.
<https://doi.org/10.48550/arXiv.2509.04664>
30. Karpatne, A., Deshwal, A., Jia, X., Ding, W., Steinbach, M., Zhang, A., Reddy, C.K., Roy, S., Shen, Y., Wang, Y., Wu, X., Xie, Y., Yin, M., Zhao, Y., Anandkumar, A., Banerjee, A., Bengio, Y., Chawla, N. V., Chen, C., ... Kumar, V. (2025). AI-enabled scientific revolution in the age of generative AI: Second NSF workshop report. *npj Artificial Intelligence*, 1, Article 18. <https://doi.org/10.1038/s44387-025-00018-6>
31. Krajowa Izba Radców Prawnych (2025) AI in the work of an attorney-at-law: Recommendations on how attorneys-at-law should use AI-based tools (English translation by Sołtysiński Kawecki &

- Szłęczak – Kancelaria Radców Prawnych i Adwokatów Sp. K.). Krajowa Izba Radców Prawnych.
<https://kirp.pl/wp-content/uploads/2025/05/rekomendacje-ENG-NET.pdf>
32. Lai J, Gan W, Wu J, Qi Z, Yu PS (2024) Large language models in law: A survey. *AI Open* 5:181–196.
<https://doi.org/10.1016/j.aiopen.2024.09.002>
 33. Leopold M (2024) AI in education: Promise, peril and a path forward. *Legal Inform Manage* 24(4):230–236. <https://doi.org/10.1017/S1472669624000513>
 34. Lorteau S, Sarro D (2026) Artificial intelligence in legal education: A scoping review. *Law Teacher* 60(1):138–162
 35. Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z (2023) ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *J Association Inform Sci Technol* 74(5):570–581.
<https://doi.org/10.1002/asi.24750>
 36. *Mata v, Avianca (2023) Inc.*, 678 F. Supp. 3d 443S.D.N.Y
 37. Matthee J, Stopforth G (2025) The importance of ethical and responsible AI training in law schools to avoid career-ending pitfalls for aspiring attorneys. *J Juridical Sci* 50(1):49–67.
<https://doi.org/10.38140/jjs.v50i1.9302>
 38. Mogbojuri BO (2024), June 4 Generative AI: Definition, concepts, applications, and future prospects. *TechRxiv*. <https://doi.org/10.36227/techrxiv.171746875.59016695/v1>
 39. Murray MD (2025) Algorithmic ethics in an era of agentic AI advocacy: An analysis of AI's impact on the Model Rules of Professional Conduct and the Model Code of Judicial Conduct [Preprint].
<https://doi.org/10.2139/ssrn.5510560>. SSRN
 40. Needham CA (2025) Regulation of the use of generative artificial intelligence tools in the delivery of legal services: Verification and accountability. *Wash Univ J Law Policy* 77(1):184–211
 41. OpenAI (2024) Video generation models as world simulators. <https://openai.com/index/video-generation-models-as-world-simulators/>
 42. Oxford Institute (2025) Brazil: Global leader in AI adoption in criminal justice.
<https://www.techandjustice.bsg.ox.ac.uk/research/brazil>
 43. Pangandaman H, Hayudini MA, Tandih F, Jupakkal M, Abdurasul JN, Najar A, Mercado C, Amang DDI (2025) Exploring the role of large language models (LLMs) as an academic resource for students: A scoping review
 44. Pedró F, Mendigutxia A (2025) The role of higher education in national artificial intelligence strategies: A comparative policy review. UNESCO IESALC
 45. Pimentel Bernal PM, Veliz JC (2025) Gender Bias in Generative Artificial Intelligence: A Literature Review on Perspectives and Implications. *International Conference on Gender Research*, 8(1), 508–511. <https://doi.org/10.34190/icgr.8.1.3402>
 46. Resühr D, Garnett C (2025) The good, the bad, and the ugly of AI in medical imaging. *EMJ Radiol* 6(1):53–55. <https://doi.org/10.33590/emjradiol/jtpo7801>

47. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2021) High-Resolution Image Synthesis with Latent Diffusion Models. arXiv. <https://doi.org/10.48550/arXiv.2112.10752>
48. Sanz-Tejeda A, Domínguez-Oller JC, Baldaquí-Escandell JM, Gómez-Díaz R, García-Rodríguez A (2026) The impact of generative AI on academic reading and writing: A synthesis of recent evidence (2023–2025). *Frontiers in Education*, 10, Article 1711718
<https://doi.org/10.3389/feduc.2025.1711718>
49. Terzidou K (2025) Generative AI systems in legal practice offering quality legal services while upholding legal ethics. *Int J Law Context* 21:431–452
50. UNESCO (2025) Shaping the future: Higher education's impact on national AI policies
51. UNESCO & IRCAI (2024) Challenging systematic prejudices: An investigation into gender bias in large language models
52. UN Women, & Everyday Sexism Project (2025) How AI is exacerbating technology-facilitated violence against women and girls
53. Universidad de Buenos Aires IALAB (2023) Guía de directrices para el uso de ChatGPT e IA generativa de texto en la Justicia
54. University of Cambridge (2026) AI guidance
55. Verspoor K, Cohen KB (2013) Natural language processing. *Encyclopedia Syst Biology* 1495–1498.
https://doi.org/10.1007/978-1-4419-9863-7_158
56. Vicci H (2024) Emotional intelligence in artificial intelligence: A review and evaluation study.
<https://doi.org/10.2139/ssrn.4818285>. SSRN
57. Weinmann H, Messingschlager TV, Appel M (2026) Gender bias in text-to-image generative artificial intelligence: Neglect and stereotypical presentations across three popular platforms. *New Media & Society*. <https://doi.org/10.1177/14614448261435197>
58. Wiesinger J, Marlow P, Vuskovic V (2024), septiembre Agents [White paper]. Google. Recuperado de https://ppc.land/content/files/2025/01/Newwhitepaper_Agents2.pdf
59. World Economic Forum, & Capgemini (2024), December 16 Navigating the AI frontier: A primer on the evolution and impact of AI agents [Report]. World Economic Forum.
https://reports.weforum.org/docs/WEF_Navigating_the_AI_Frontier_2024.pdf
60. Zhai C, Wibowo S, Li LD (2024) The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learn Environ* 11., Article 28.
<https://doi.org/10.1186/s40561-024-00316-7>
61. Zhang C, Zhang C, Zhang M, Kweon IS (2023) Text-to-image diffusion models in generative AI: A survey. arXiv. Hemerographic references
62. Oxford Institute (2025) *Brazil | Global leader in AI adoption in criminal justice*. Ox.ac.uk.
<https://www.techandjustice.bsg.ox.ac.uk/research/brazil>
63. Cohen T (2025) *Darrow · 10 Best AI Tools for Lawyers in 2025*. Darrow.ai.
<https://www.darrow.ai/resources/ai-tools-for-lawyers>

64. Allawyer (2026) *AI Lawyer | Your personal AI legal assistant*. Linguana.aillawyer.pro.
<https://aillawyer.pro/>
65. Dunphy K (2026) *Spellbook*. Spellbook.legal. <https://www.spellbook.legal/learn/legal-ai-tools>
66. UNESCO (2025b), December 3 *AI in the Courtroom: UNESCO's New Guidelines for the Judiciary*.
Unesco.org. <https://www.unesco.org/en/articles/ai-courtroom-unescos-new-guidelines-judiciary>
67. Merken S (2025), May 23 *Trouble with AI hallucinations spreads to big law firms*. Reuters.
<https://www.reuters.com/legal/government/trouble-with-ai-hallucinations-spreads-big-law-firms-2025-05-23/>
68. Schlissel E, Caffarone D (2026), March 10 *Loose AI Prompts Sink Ships: How Heppner Shook the Legal Community - New York State Bar Association*. New York State Bar Association - NYSBA.
<https://nysba.org/loose-ai-prompts-sink-ships-how-heppner-shook-the-legal-community/>
69. Roberts O (2026), February 20 *Spread of AI Hallucinations Drives Need for Sanctions Reporting*.
@BLaw. <https://news.bloomberglaw.com/legal-exchange-insights-and-commentary/spread-of-ai-hallucinations-drives-needfor-sanctions-reporting>
70. Lee M (2026), February 8 *AI Hallucinations in Court: 38UK Sanctions & Case Law Update*. Natural and Artificial Law. <https://naturalandartificiallaw.com/ai-hallucinations-in-court-cases-38uk/>
71. Agraz A (2026), March 6 *Google Sued Over Gemini AI Chatbot's Alleged Role in Man's Suicide*.
Lawcommentary.com. <https://www.lawcommentary.com/articles/google-sued-over-gemini-ai-chatbots-alleged-role-in-mans-suicide>
72. Goyal P (2025), April 29 *AI in Court: When Legal Tech Goes Rogue – Lessons from Mata v. Avianca* /
Virtuosity Legal. Virtuosity Legal | Substance. Authority. Insight. <https://virtuositylegal.com/ai-in-court-when-legal-tech-goes-rogue-lessons-from-mata-v-avianca/>
73. Katariya N (2025), October 20 *GenAI vs LLM: What's the Real Difference?* | Opensense Labs.
Opensenselabs.com. <https://opensenselabs.com/blog/genai-vs-llm>
74. Mitchell M (2026) March 5). *No, AI is not a Stochastic Parrot* 🦜. Medium.
<https://medium.com/@margarmitchell/no-ai-is-not-a-stochastic-parrot-a99e57766bed>