Guía Práctica para el Cumplimiento del IA Act Abogados

derechoartificial.com

Índice

• **4.4. Sinergia IA Act - RGPD:** Protección de datos desde el diseño y minimización en el entrenamiento de modelos.

Capítulo 5. Mitigación Técnica: RAG y Arquitecturas de IA Fiable

• **5.1. Generación Aumentada por Recuperación (RAG):** Anclaje de respuestas en fuentes legales autorizadas y vigentes.
• **5.2. Jerarquía normativa en agentes de IA:** Implementación de la pirámide de Kelsen para evitar alucinaciones de invalidez.
• **5.3. Explicabilidad e Interpretabilidad (XAI):** De la fundamentación a la interpretación razonada de resultados.
• **5.4. Verificación post-hoc:** Módulos automatizados de chequeo factual y calibración de confianza.

Capítulo 6. Gobernanza y Plan de Acción: Hacia un Despacho "Safe-by-Design"

• **6.1. Protocolos internos de control:** Definición de herramientas autorizadas, registros de uso y responsables de supervisión.
• **6.2. El principio de "Reserva de Humanidad":** Mandato de revisión humana universal de toda decisión relevante.
• **6.3. Alfabetización y capacitación continua:** Planes de formación obligatoria para el personal del despacho.
• **6.4. Auditoría y rendición de cuentas:** Mantenimiento de la documentación técnica y vigilancia poscomercialización.

CAPÍTULO 1. FUNDAMENTOS: EL DESAFÍO DE LA VERACIDAD EN LA ABOGACÍA DIGITAL

1.1. Propósito y alcance de la guía: de la curiosidad tecnológica a la responsabilidad profesional

La integración de la inteligencia artificial (IA) en la práctica jurídica ha trascendido la fase de experimentación técnica para convertirse en un imperativo de **responsabilidad profesional** y diligencia debida. El propósito de esta guía es establecer un marco normativo y operativo que permita a los letrados transitar desde la mera observación de la tecnología hacia una adopción estratégica basada en la **seguridad jurídica**. Se prescribe que la competencia tecnológica no es ya una facultad opcional, sino una exigencia ineludible para el ejercicio profesional moderno. El alcance de esta norma técnica abarca la totalidad del ciclo de vida de los sistemas de IA empleados en el despacho, desde su selección y entrenamiento hasta su supervisión y auditoría postcomercialización. La abogacía, como garante de la **tutela judicial efectiva**, debe liderar la transformación digital asegurando que la innovación no debilite, sino que refuerce los pilares del Derecho.

1.2. Definiciones clave según el IA Act: sistemas de IA, modelos de uso general y capacidad de inferencia

Para la correcta subsunción de las herramientas tecnológicas en el marco legal vigente, es obligatorio adherirse a la taxonomía establecida en el Reglamento (UE) 2024/1689 (IA Act):

- **Sistema de IA:** Se define como un sistema basado en una máquina diseñado para funcionar con **distintos niveles de autonomía**, capaz de mostrar adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada la manera de generar resultados de salida.

- **Capacidad de Inferencia:** Representa la facultad técnica del sistema para deducir modelos o algoritmos a partir de datos, permitiendo el aprendizaje, el razonamiento o la modelización fáctica.

- **Modelos de IA de uso general (GPAI):** Son arquitecturas de IA que presentan un grado significativo de **generalidad**, capaces de realizar de manera competente una amplia variedad de tareas diferenciadas y que pueden integrarse en múltiples sistemas posteriores.

- **Responsable del despliegue (*Deployer*):** En el contexto de un despacho de abogados, esta figura recae sobre la persona física o jurídica que utiliza un sistema de IA bajo su propia autoridad en el ejercicio de su actividad profesional.

1.3. La dicotomía fundamental: IA Generativa (el oráculo creativo) frente a IA Consultiva (el archivero experto)

La arquitectura técnica de la herramienta determina su nivel de riesgo y su aplicabilidad legal. Se establece una distinción prescriptiva entre dos paradigmas:

1. **IA Generativa (Propósito General):** Operan mediante la predicción probabilística del siguiente token en una secuencia, optimizando la fluidez y la coherencia conversacional sobre la veracidad fáctica. Estos sistemas son inherentemente propensos a las **alucinaciones** —generación de información plausible pero falsa— debido a que su diseño prioriza evitar el silencio en lugar de admitir el desconocimiento. Su funcionamiento se caracteriza por ser un **algoritmo opaco** o "caja negra", donde es imposible trazar el origen de una afirmación específica.

2. **IA Consultiva (Especializada):** Basada en arquitecturas de **Generación Aumentada por Recuperación (RAG)**, este modelo no crea conocimiento, sino que lo recupera de un corpus externo, curado y autorizado. Su objetivo es la **fundamentación** (*grounding*), actuando como un asistente experto que cita sus fuentes de manera transparente, permitiendo la trazabilidad y la verificación por parte del profesional humano.

1.4. Metodología de adaptación estratégica: delimitación de herramientas y escenario de adopción

La metodología para la adaptación de un despacho al IA Act requiere una evaluación multidimensional que contemple los **niveles de riesgo** y el impacto en los **derechos fundamentales**. En España, el escenario actual muestra una implementación moderada pero acelerada, donde el 60% de los despachos que han adoptado la IA lo han hecho en el último año.

Se prescribe una **metodología de adaptación** basada en cuatro ejes:

1. **Clasificación de Riesgos:** Identificar si las herramientas empleadas se subsumen en las categorías de "Alto Riesgo" (especialmente aquellas destinadas a la administración de justicia o procesos de selección) o en las de "Riesgo Limitado" con obligaciones de transparencia.

2. **Principio de No Sustitución:** La tecnología debe ser considerada exclusivamente como un complemento instrumental; la toma de decisiones legales críticas es una actividad humana irreductible e indelegable.

3. **Gestión de la Opacidad:** Implementar medidas técnicas que mitiguen el efecto "caja negra", exigiendo a los proveedores niveles adecuados de explicabilidad y auditabilidad.

4. **Alfabetización en IA:** El despacho tiene la obligación legal de garantizar que su personal posea un nivel suficiente de conocimientos para comprender los riesgos y limitaciones de los sistemas utilizados.

CAPÍTULO 2. EL ECOSISTEMA TECNOLÓGICO: CLASIFICACIÓN Y CICLO DE VIDA

2.1. Funcionalidades de la IA en el despacho: de la revisión documental a la analítica predictiva

La implementación de la inteligencia artificial (IA) en la oficina de farmacia jurídica no es monolítica, sino que se bifurca en dos categorías funcionales prescritas por la doctrina técnica: la **IA analítica/predictiva** y la **IA generativa**. Se prescribe que el responsable del despliegue identifique la naturaleza de la tarea para mitigar riesgos de precisión. La IA analítica se especializa en la detección de patrones en grandes bases de datos legales, la clasificación de información y el soporte a decisiones mediante la evaluación de riesgos basada en precedentes y estadísticas. Por otro lado, la IA generativa se orienta a la creación de contenido nuevo (escritos, contratos, resúmenes) imitando estructuras lingüísticas aprendidas. La capacidad de procesamiento de estos sistemas supera considerablemente las posibilidades humanas en tareas de elevada carga burocrática y repetitiva, liberando recursos para funciones donde el juicio profesional es insustituible.

2.2. Mapa de herramientas relevantes: análisis de soluciones sectoriales

El mercado de **LegalTech** actual ofrece un ecosistema diversificado de herramientas que el despacho debe auditar bajo criterios de diligencia reforzada. Se clasifican según su funcionalidad crítica:

- **Analítica y Predictiva:** Soluciones como *vLex Analytics*, *Jurimetría* y *Tirant Analytics* permiten identificar tendencias judiciales y predecir resultados procesales mediante el análisis de patrones en sentencias.

- **Gestión y Transcripción:** Herramientas como *Copilot*, *Fireflies.ai* y *DigalawX* automatizan la conversión de voz a texto estructurado en vistas y reuniones.

- **Investigación y Redacción Asistida:** Plataformas especializadas como *Vincent AI (vLex)*, *Iberley IA*, *CoCounsel (Thomson Reuters)* y *Justicio* integran bases de datos verificadas para fundamentar respuestas legales y generar borradores.

- **IA de Propósito General:** Modelos como *ChatGPT* y *Claude* ofrecen versatilidad en redacción y síntesis, pero presentan niveles de riesgo elevados de alucinación si no se integran con arquitecturas de recuperación externa.

2.3. Ciclo de vida de una solución de IA legal: fases operativas y tratamiento de datos

Se prescribe que todo sistema de IA integrado en un despacho sea evaluado conforme a un ciclo de vida normativo que consta de cuatro etapas críticas:

1. **Concepción y análisis:** Fijación de requisitos funcionales y restricciones normativas derivados del mercado jurídico.

2. **Desarrollo:** Fase que incluye el entrenamiento mediante algoritmos de **Machine Learning** (ML), pruebas de validación y verificación de la exactitud de los datos. Si se utilizan datos de personas físicas en el entrenamiento, esta fase constituye un tratamiento sujeto al RGPD.

3. **Explotación:** Comprende el despliegue, la inferencia (generación de resultados) y el mantenimiento evolutivo del sistema. Es obligatorio realizar recalibraciones periódicas cuando el modelo se realimenta con interacciones del interesado.

4. **Retirada:** Descarte del componente por obsolescencia o decisión del responsable, lo que exige protocolos de supresión de datos locales o distribuidos.

2.4. Integración técnica: Modelos de Lenguaje a Gran Escala (LLM) y el rol del Deep Learning

La arquitectura técnica predominante en la abogacía digital se sustenta en el **Deep Learning** (aprendizaje profundo), que emula redes neuronales artificiales de múltiples capas para procesar información compleja. Los **Large Language Models (LLM)** son algoritmos diseñados para procesar y generar texto prediciendo probabilísticamente el siguiente *token* (unidad de procesamiento) en una secuencia.

Se establece una distinción técnica prescriptiva entre el conocimiento **paramétrico** (estático e interno del modelo) y el conocimiento **externo** (dinámico y recuperado). La integración técnica más fiable para el sector legal es la **Generación Aumentada por Recuperación (RAG)**. Esta arquitectura funciona en dos fases: primero, un módulo de recuperación identifica fragmentos relevantes en un corpus legal curado (Vectorstore) mediante búsqueda de similitud semántica; segundo, estos datos aumentan el *prompt* inyectado al LLM para asegurar que la respuesta esté estrictamente fundamentada en fuentes autorizadas y vigentes.

CAPÍTULO 3. ANÁLISIS MULTIDIMENSIONAL DE RIESGOS E IMPACTO DEONTOLÓGICO

3.1. Riesgos sistémicos: opacidad ("caja negra"), sesgos algorítmicos y el fenómeno de las alucinaciones

El despliegue de sistemas de inteligencia artificial (IA) en la abogacía introduce riesgos sistémicos que desafían la integridad de la práctica jurídica. Se prescribe, en primer término, la gestión de los **algoritmos opacos** o sistemas de "caja negra", definidos como arquitecturas donde la lógica subyacente y el proceso de decisión no resultan comprensibles para el ser humano. Esta opacidad compromete la trazabilidad exigida por el **IA Act** y dificulta la justificación de resultados ante clientes y tribunales.

En segundo término, la doctrina técnica identifica las **alucinaciones** como una patología endémica de los modelos de lenguaje a gran escala (LLM). Estas consisten en la generación de información que, siendo lingüísticamente coherente y plausible, carece de veracidad fáctica o fundamento legal. Se clasifican de forma prescriptiva en:

1. **Factuales/extrínsecas:** fabricación de autoridades o precedentes inexistentes.
2. **De fundamentación (*misgrounding*):** citación de fuentes reales pero con una interpretación que estas no respaldan o incluso contradicen.

Finalmente, el **sesgo algorítmico** constituye un riesgo de vulneración de **derechos fundamentales**, derivado de datos de entrenamiento no representativos o que codifican prejuicios históricos. Es imperativo vigilar el **sesgo de automatización**, que induce al profesional a confiar acríticamente en los resultados del sistema, reduciendo su capacidad de cuestionamiento.

3.2. Precedentes judiciales de negligencia: lecciones del caso *Mata v. Avianca* y la sanción del Tribunal Constitucional español

La jurisprudencia reciente ha establecido límites claros a la delegación de tareas en la IA. Se establecen como precedentes obligatorios de estudio:

- **Caso** *Mata v. Avianca* **(S.D.N.Y. 2023):** Sanción impuesta a letrados por presentar un escrito con seis citas judiciales inexistentes fabricadas por un chatbot. La lección prescriptiva es que la pregunta a la propia IA sobre su veracidad es manifiestamente insuficiente.

- **Sanción del Tribunal Constitucional (España, Nota 90/2024):** El tribunal sancionó por unanimidad a un abogado que incluyó 19 citas de doctrina constitucional inexistentes bajo el pretexto de una "desconfiguración de la base de datos". El tribunal prescribe que la **responsabilidad del letrado es absoluta e independiente de la herramienta**, manteniendo la obligación de revisar exhaustivamente todo contenido antes de su envío procesal.

3.3. El deber de competencia tecnológica: la obligación ética de comprender las limitaciones de la IA

Se establece la **Obligación de Competencia Tecnológica** como un nuevo pilar deontológico. El letrado tiene el deber de poseer una alfabetización suficiente en IA que le permita evaluar críticamente los riesgos de los sistemas empleados. La diligencia profesional exige no solo un análisis *ex ante* de la fiabilidad del sistema, sino un examen *ex post* obligatorio de cada resultado. Este deber se fundamenta en el **principio de reserva de humanidad**, el cual prescribe que el juicio profesional es irreductible e insustituible por la automatización.

3.4. Secreto profesional y privacidad: riesgos derivados del *cloud computing* y fugas de datos

El uso de sistemas de IA basados en infraestructuras de **nube (***cloud computing***)** introduce riesgos críticos para el **secreto profesional** y la protección de datos (RGPD). Se prescriben los siguientes riesgos específicos:

- **Acceso gubernamental externo:** La normativa extranjera, como la *Cloud Act* de EE. UU., puede permitir el acceso a datos almacenados en servidores fuera de la jurisdicción nacional o europea.

- **Fuga de datos (***data leaks***):** Riesgo de que información sensible del cliente sea reutilizada por el modelo para generar respuestas a otros usuarios.

- **Seguridad de la información:** Es obligatorio realizar evaluaciones de impacto (EIPD) antes del despliegue de sistemas de IA que traten datos de personas físicas para garantizar la integridad y confidencialidad exigida por el Artículo 5 del RGPD.

CAPÍTULO 4. MARCO REGULATORIO: EL IA ACT COMO NORMA DE SEGURIDAD DE PRODUCTO

4.1. Clasificación de riesgos según el IA Act: de sistemas prohibidos a sistemas de riesgo mínimo

El Reglamento (UE) 2024/1689 (IA Act) establece un marco jurídico uniforme basado en un **enfoque estratificado de riesgos**, diseñado para equilibrar la innovación con la protección de la salud, la seguridad y los derechos fundamentales. Se prescribe imperativamente la clasificación de los sistemas de IA en cuatro niveles normativos:

1. **Riesgo Inaceptable (Prácticas Prohibidas):** El Artículo 5 prohíbe taxativamente los sistemas que empleen técnicas subliminales o deliberadamente manipuladoras que alteren el comportamiento humano mermando la capacidad de decisión informada. Asimismo, se prohíbe la puntuación ciudadana (*social scoring*) basada en comportamiento social o características personales que deriven en tratos desfavorables injustificados.

2. **Alto Riesgo:** Sistemas sujetos a requisitos obligatorios de seguridad y evaluación de conformidad previa a su introducción en el mercado. Incluyen componentes de seguridad de productos regulados y sistemas específicos enumerados en el Anexo III.

3. **Riesgo Limitado:** Sistemas sujetos exclusivamente a obligaciones de transparencia, como los destinados a interactuar con personas físicas o generar contenidos sintéticos (ultrasuplantaciones o *deepfakes*), los cuales deben ser etiquetados de forma clara y distinguible.

4. **Riesgo Mínimo o Nulo:** Sistemas que no presentan riesgos significativos para los derechos o la seguridad, los cuales no están sujetos a obligaciones específicas más allá del cumplimiento general de la normativa vigente.

4.2. La abogacía ante los sistemas de "Alto Riesgo": impacto en la administración de justicia y procesos democráticos

Dentro de la taxonomía del IA Act, la práctica jurídica se ve directamente interpelada por la categoría de **Alto Riesgo** definida en el Anexo III, punto 8. Se establece que los sistemas de IA destinados a ser utilizados por una autoridad judicial, o en su nombre, para asistir en la investigación e interpretación de hechos y del Derecho, así como en la aplicación de la ley a casos concretos, poseen este nivel de riesgo debido a su potencial impacto en la **tutela judicial efectiva** y un juez imparcial.

La norma prescribe que el uso de estas herramientas debe limitarse a funciones de apoyo, estableciendo que **la toma de decisiones finales es una actividad humana irreductible** que no puede ser sustituida por la IA. Las herramientas analíticas de despachos que pretendan influir en la estrategia judicial o predecir resultados procesales deben ser auditadas bajo este estándar, asegurando que su opacidad no limite el derecho de defensa ni la capacidad de impugnar decisiones ante los tribunales.

4.3. Roles y responsabilidades: el despacho como "responsable del despliegue" (*deployer*) frente al "proveedor"

Es preceptivo que el despacho identifique su rol jurídico conforme al Artículo 3 para determinar su régimen de responsabilidad:

- **Responsable del despliegue (*Deployer*):** Persona física o jurídica que utiliza un sistema de IA bajo su propia autoridad en su actividad profesional. La mayoría de los despachos operan bajo este rol [160, 160 nota 2]. Sus obligaciones incluyen: (i) adoptar medidas técnicas y organizativas para el uso del sistema según las instrucciones del proveedor; (ii) garantizar la supervisión humana por personas cualificadas y con autoridad; y (iii) vigilar el funcionamiento para detectar posibles riesgos o incidentes graves.

- **Proveedor (*Provider*):** Quien desarrolla un sistema de IA o un modelo de IA de uso general con el fin de introducirlo en el mercado bajo su propio nombre o marca. Un despacho será considerado proveedor si modifica sustancialmente un sistema de IA de alto riesgo o si cambia su finalidad prevista.

El responsable del despliegue asume una **responsabilidad subjetiva por falta de diligencia** en la supervisión de los resultados, siendo responsable de validar y, en su caso, corregir las alucinaciones o errores del sistema antes de su incorporación al servicio jurídico.

4.4. Sinergia IA Act - RGPD: protección de datos desde el diseño y minimización en el entrenamiento de modelos

El IA Act no deroga, sino que complementa el Reglamento General de Protección de Datos (RGPD). Se prescribe una integración operativa de ambas normas a través de los siguientes ejes:

1. **Evaluación de Impacto (EIPD):** Los responsables del despliegue deben utilizar la documentación técnica facilitada por el proveedor (bajo el Art. 13 del IA Act) para dar cumplimiento a la obligación de realizar evaluaciones de impacto relativas a la protección de datos exigidas por el Art. 35 del RGPD.

2. **Calidad y Gobernanza de Datos:** El Artículo 10 del IA Act exige que los conjuntos de datos de entrenamiento sean pertinentes, representativos y libres de errores. Se permite excepcionalmente el tratamiento de **categorías especiales de datos** (Art. 9 RGPD) únicamente cuando sea estrictamente necesario para detectar y corregir sesgos algorítmicos, siempre que se apliquen salvaguardias reforzadas y se evite la reidentificación de los interesados.

3. **Transparencia y Derechos del Interesado:** Se garantiza el derecho de las personas afectadas a obtener una **explicación clara y significativa** sobre el papel de la IA en decisiones automatizadas que produzcan efectos jurídicos o les afecten significativamente. La transparencia debe permitir al profesional y al cliente comprender la lógica aplicada y las limitaciones de exactitud del sistema.

CAPÍTULO 5. MITIGACIÓN TÉCNICA: RAG Y ARQUITECTURAS DE IA FIABLE

5.1. Generación Aumentada por Recuperación (RAG): anclaje de respuestas en fuentes legales autorizadas y vigentes

Se prescribe el abandono del paradigma de la IA Generativa de propósito general —el "oráculo creativo"— en favor de la **IA Consultiva especializada** o "archivero experto" para tareas que exijan veracidad fáctica. Mientras que los sistemas generativos operan en un modo de "libro cerrado" basado en conocimiento paramétrico estático y probabilístico, la arquitectura de **Generación Aumentada por Recuperación (RAG)** transforma el sistema en un modelo de "libro abierto".

El mecanismo operativo de RAG debe ejecutar una secuencia técnica obligatoria: 1) **Recuperación (*Retrieval*):** identificación y extracción de fragmentos (*chunks*) de un corpus documental externo, curado y autorizado; y 2) **Generación:** aumento del *prompt* del usuario con el contexto recuperado para que el modelo genere una respuesta **estrictamente fundamentada**. Se establece que la efectividad de RAG no es absoluta; estudios empíricos demuestran que, aunque reduce las alucinaciones frente a modelos base, las herramientas comerciales mantienen tasas de error residuales de entre el 17% y el 34%, lo que exige una optimización continua de la base de conocimiento y del módulo de recuperación.

5.2. Jerarquía normativa en agentes de IA: implementación de la pirámide de Kelsen para evitar alucinaciones de invalidez

La arquitectura de los **agentes de IA legales** debe integrar de forma nativa la **jerarquía normativa de Hans Kelsen** para garantizar la validez legal de sus inferencias. Se prescribe que un sistema carente de conciencia jerárquica es inherentemente propenso a la "alucinación de invalidez", al poder fundamentar respuestas en normas de rango inferior (como reglamentos municipales) que han sido invalidadas por normas superiores (leyes o la Constitución).

Para mitigar este riesgo, las arquitecturas fiables deben implementar:

1. **Bases de conocimiento estructuradas:** etiquetado de documentos con metadatos que reflejen su rango jerárquico (Constitución, Ley Orgánica, etc.).

2. **Algoritmos de recuperación sensibles a la jerarquía:** priorización sistemática de las fuentes de mayor autoridad en la fase de *retrieval*.

3. **Módulos de validación kelseniana:** verificación obligatoria de que toda interpretación propuesta es coherente con las normas de rango superior y el estado de vigencia actual.

5.3. Explicabilidad e Interpretabilidad (XAI): de la fundamentación a la interpretación razonada de resultados

Frente al fenómeno de los **algoritmos opacos** o "cajas negras", se prescribe la adopción de la **Inteligencia Artificial Explicable (XAI)** como requisito previo para la tutela judicial efectiva y el respeto a los derechos fundamentales. El desarrollo de la XAI jurídica debe transitar por tres niveles progresivos de madurez técnica:

- **Respuestas Fundamentadas:** capacidad de anclar y citar cada afirmación en una fuente verificable (trazabilidad básica).

- **Respuestas Argumentadas:** externalización de los pasos lógicos de la inferencia, demostrando una cadena de razonamiento coherente (explicabilidad del proceso).

- **Interpretación Razonada:** nivel avanzado donde la IA justifica por qué opta por una interpretación específica frente a otras alternativas plausibles, ponderando matices y ambigüedades (explicabilidad sustantiva).

5.4. Verificación post-hoc: módulos automatizados de chequeo factual y calibración de confianza

Dada la imposibilidad técnica de eliminar totalmente las alucinaciones en la fase de generación, es preceptivo implementar una "última línea de defensa" mediante **mecanismos de verificación post-hoc**. Estos módulos deben realizar una comprobación factual automatizada (*Fact-Checking*) comparando las afirmaciones generadas contra bases de conocimiento canónicas y bases de datos jurisprudenciales con metadatos de derogación.

Asimismo, el sistema debe incorporar de forma obligatoria:

- **Validación sintáctica y lógica:** uso de heurísticas deterministas para verificar el formato de las citas y la coherencia temporal de las sentencias mencionadas.

- **Comunicación transparente de la incertidumbre:** los sistemas deben presentar puntuaciones de confianza calibradas y emplear el **"silencio estratégico"** (abstención justificada) cuando la información recuperada sea ambigua o insuficiente, evitando la propensión al sesgo de automatización.

- **Citación a nivel de fragmento:** vinculación inequívoca de cada conclusión al pasaje exacto de la fuente que la respalda para facilitar la supervisión humana experta.

CAPÍTULO 6. GOBERNANZA Y PLAN DE ACCIÓN: HACIA UN DESPACHO "SAFE-BY-DESIGN"

6.1. Protocolos internos de control: definición de herramientas autorizadas, registros de uso y responsables de supervisión

Se prescribe de forma imperativa que los despachos de abogados, en su calidad de **responsables del despliegue** (*deployers*), implementen un modelo de gobernanza de la información efectivo basado en el principio de **responsabilidad proactiva** o *accountability*. Es obligatorio que el despacho adopte medidas técnicas y organizativas apropiadas para garantizar y demostrar que el tratamiento de datos y el uso de sistemas de IA son conformes con la legalidad vigente.

Para ello, resulta preceptivo el establecimiento de **protocolos internos de control** que deben: (i) definir taxativamente las herramientas tecnológicas autorizadas y sus finalidades específicas; (ii) designar responsables de supervisión con autoridad y competencia técnica; y (iii) garantizar la trazabilidad mediante el registro documental de cada decisión relevante asistida por IA. El registro debe incluir detalladamente los algoritmos empleados, los datos de entrenamiento (cuando el despacho actúe como proveedor) y los criterios de evaluación utilizados para validar los resultados.

6.2. El principio de "Reserva de Humanidad": mandato de revisión humana universal de toda decisión relevante

La supervisión humana se erige como el eje vertebrador de la integración de la IA en la abogacía, constituyendo una salvaguarda indispensable para prevenir o reducir al mínimo los riesgos para los **derechos fundamentales**. Se prescribe el **Mandato de Revisión Humana Universal**: toda información o borrador generado por un sistema de IA debe ser revisado, validado y, en su caso, corregido por un profesional de la abogacía antes de su incorporación al servicio jurídico o su presentación ante tribunales.

El principio de **"Reserva de Humanidad"** establece que la toma de decisiones legales finales es una actividad humana irreductible e insustituible; la tecnología es una herramienta de apoyo, pero no puede reemplazar el juicio profesional del letrado. La jurisprudencia constitucional prescribe que la responsabilidad del letrado es **absoluta e independiente** de la herramienta, manteniendo la obligación de revisar exhaustivamente todo contenido antes de su envío procesal. En consecuencia, queda prohibido delegar en la IA la parte esencial del trabajo intelectual o la toma de decisiones jurídicas materiales.

6.3. Alfabetización y capacitación continua: planes de formación obligatoria para el personal del despacho

La **alfabetización en materia de IA** ha dejado de ser una recomendación técnica para convertirse en una **obligación legal** establecida en el Artículo 4 del Reglamento (UE) 2024/1689. Los responsables del despliegue deben adoptar medidas para garantizar que su personal posea un nivel suficiente de conocimientos, teniendo en cuenta sus funciones y el contexto de uso de los sistemas.

Se prescribe la implementación de planes de **formación continua y actualizada** que permitan al personal: (i) comprender el funcionamiento básico y las capacidades de los sistemas empleados; (ii) identificar riesgos sistémicos como las **alucinaciones** y los sesgos algorítmicos; y (iii) interpretar correctamente los resultados de salida para evitar el **sesgo de automatización**. La competencia tecnológica es una exigencia ineludible para el ejercicio profesional, orientada a promover una cultura de aprendizaje que refuerce la calidad del servicio jurídico.

6.4. Auditoría y rendición de cuentas: mantenimiento de la documentación técnica y vigilancia poscomercialización

El régimen de rendición de cuentas exige que los despachos mantengan una documentación exhaustiva de sus procesos y evaluaciones vinculadas al uso de la IA. Es preceptivo conservar la **documentación técnica** y, cuando proceda, los certificados de conformidad durante un período de **diez años** a contar desde la puesta en servicio del sistema.

Asimismo, se prescribe la implementación de un sistema de **vigilancia poscomercialización**. Este sistema debe recopilar y analizar de manera activa datos sobre el funcionamiento real de las herramientas para detectar posibles riesgos, incidentes graves o fallos de precisión de forma oportuna. Las auditorías deben ser dinámicas, evaluando las divergencias entre los resultados previstos y los obtenidos, para adoptar medidas correctivas inmediatas, incluyendo la supresión de información o la desactivación del sistema si se vulneran estándares de seguridad o derechos de los interesados.

A continuación, se presenta el listado de referencias bibliográficas de las fuentes utilizadas en el análisis y redacción de los capítulos precedentes, siguiendo las normas de la **APA 7ª Edición**:

Agencia Española de Protección de Datos. (2020, febrero). *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción.*.

Agencia Española de Supervisión de Inteligencia Artificial. (2025). *Código de Ética Institucional. Normas internas de integridad, ética y evaluación de riesgos.*.

Ayo Ferrándiz, C., Seijo Bar, Á., Garre Anguera de Sojo, I., & González Guillén, P. (2025). Responsabilidad civil e inteligencia artificial. *Actualidad Jurídica Uría Menéndez*, (67), 29-56..

Consejo General de la Abogacía Española & Ilustre Colegio de Abogados de Valencia. (2025). *Libro Blanco sobre Inteligencia Artificial y Abogacía.*.

Dantart, A. (2025). *Inteligencia artificial jurídica y el desafío de la veracidad: Análisis de alucinaciones, optimización de RAG y principios para una integración responsable* [Informe técnico]. arXiv..

European Central Bank. (2024). *ESCB Legal Conference 2024.*.

Ilustre Colegio de la Abogacía de Madrid. (2025). *Guía ICAM de Buenas Prácticas para el uso de la Inteligencia Artificial (IA) en la Abogacía.*.

Parlamento Europeo y Consejo de la Unión Europea. (2024, 13 de junio). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.° 300/2008, (UE) n.° 167/2013, (UE) n.° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). *Diario Oficial de la Unión Europea*. http://data.europa.eu/eli/reg/2024/1689/oj.