ICAI - ICADE

GRADO EN ANÁLISIS DE NEGOCIOS / BUSINESS ANALYTICS

TRABAJO FIN DE GRADO

DESARROLLO DE UN SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA LA GESTIÓN DE INFORMACIÓN EN EL SECTOR LEGAL Y FINANCIERO

Autor: Alberto García Fábregas

Director: David Martín-Corral Calvo

MADRID | MAYO 2026

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

Desarrollo de un sistema basado en Inteligencia Artificial para la gestión de información en el sector legal y financiero.

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas

curso académico 2025/26 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Firmado por GARCIA FABREGAS ALBERTO - ***2437** el día 22/04/2026 con un certificado emitido por AC FNMT Usuarios

Fdo.: Alberto García Fábregas

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

MARTIN CORRAL CALVO DAVID - 03909768K

Digitally signed by MARTIN CORRAL CALVO DAVID - 03909768K
Date: 2026.04.22 11:50:08 +02'00'

Fdo.: David Martín-Corral Calvo

ICAI - ICADE

GRADO EN ANÁLISIS DE NEGOCIOS / BUSINESS ANALYTICS

TRABAJO FIN DE GRADO

DESARROLLO DE UN SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA LA GESTIÓN DE INFORMACIÓN EN EL SECTOR LEGAL Y FINANCIERO

Autor: Alberto García Fábregas

Director: David Martín-Corral Calvo

Madrid 2026

DESARROLLO DE UN SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA LA GESTIÓN DE INFORMACIÓN EN EL SECTOR LEGAL Y FINANCIERO

Autor: García Fábregas, Alberto.

Director: Martín-Corral Calvo, David.

Entidad Colaboradora: Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

El presente Trabajo de Fin de Grado aborda el problema de desplegar en producción un sistema de inteligencia artificial que procese documentación societaria y fiscal sensible en el sector legal, respondiendo a la brecha persistente entre la fase de prototipado y la operación productiva real. El punto de partida es la constatación de que construir un prototipo funcional que extrae información de un conjunto de documentos de muestra es técnicamente accesible con las herramientas actuales, pero convertirlo en un sistema que opera de forma continua, segura, trazable y conforme a la normativa aplicable constituye un reto de naturaleza radicalmente distinta, que involucra dimensiones técnicas, organizativas, legales y éticas que el prototipo no aborda (Sculley et al., 2015; Amershi et al., 2019).

El objetivo principal es diseñar una arquitectura de referencia auditable sobre Amazon Web Services (AWS) que garantice simultáneamente el cumplimiento del Reglamento General de Protección de Datos (Reglamento (UE) 2016/679, en adelante RGPD), del Reglamento de Inteligencia Artificial (Reglamento (UE) 2024/1689, en adelante AI Act) y de los requisitos de seguridad aplicables al caso de uso de un despacho de abogados especializado en operaciones societarias y asesoramiento fiscal que realiza análisis asistido por IA de expedientes de due diligence.

El trabajo adopta un enfoque de diseño arquitectónico basado en la revisión sistemática de literatura académica, marcos de referencia industriales (NIST AI RMF, ISO/IEC 23894, OWASP Top 10 LLM) y estándares normativos europeos, complementado con el análisis del caso de uso concreto. La solución propuesta integra una arquitectura serverless compuesta por AWS Amplify, Amazon Cognito, API Gateway, AWS Step Functions, AWS Lambda, Amazon Bedrock (con el modelo Claude de Anthropic), Amazon S3, AWS KMS, Amazon Textract, Amazon GuardDuty, Amazon CloudWatch y AWS CloudTrail, con privacidad por diseño y por defecto aplicada en cada capa del sistema. El análisis del flujo de datos de extremo a extremo identifica cinco puntos críticos de exposición y propone controles proporcionales para cada uno; el modelo de responsabilidad compartida de AWS delimita con precisión

qué aspectos de la seguridad y el cumplimiento corresponden al proveedor y cuáles al cliente, poniendo de manifiesto que contratar infraestructura gestionada no equivale a transferir la responsabilidad como responsable del tratamiento.

Los resultados muestran que el cumplimiento normativo y el diseño técnico son inseparables, que elegir un proveedor cloud de confianza no exime al despacho de sus obligaciones bajo el RGPD, y que la supervisión humana constituye el control más importante del sistema tanto desde la perspectiva de la responsabilidad profesional del abogado como desde la perspectiva del gobierno de la IA. El trabajo concluye con un itinerario de madurez estructurado en tres fases:prueba funcional con datos sintéticos, piloto controlado con datos reales acotados, e incorporación iterativa hasta la operación normalizada.

Palabras clave: Inteligencia Artificial, AWS, RGPD, AI Act, Privacidad por Diseño, Arquitectura Cloud, Sector Legal, Gobierno de la IA, Amazon Bedrock, MLOps.

ABSTRACT

This undergraduate thesis addresses the challenge of deploying in production an artificial intelligence system that processes sensitive corporate and tax documentation in the legal sector, responding to the persistent gap between the prototyping phase and real productive operation. The starting point is the observation that building a functional proof-of-concept capable of extracting information from a sample document set is technically accessible with current tools, but transforming it into a system that operates continuously, securely, traceably, and in compliance with applicable regulations represents a fundamentally different challenge—one that involves technical, organizational, legal, and ethical dimensions that the prototype does not address (Sculley et al., 2015; Amershi et al., 2019).

The main objective is to design an auditable reference architecture on Amazon Web Services (AWS) that simultaneously ensures compliance with the General Data Protection Regulation (Regulation (EU) 2016/679, hereinafter GDPR), the Artificial Intelligence Act (Regulation (EU) 2024/1689, hereinafter AI Act), and the security requirements applicable to the use case of a law firm specializing in corporate transactions and tax advisory that performs AI-assisted analysis of due diligence files.

The work adopts an architectural design approach based on the systematic review of academic literature, industrial reference frameworks (NIST AI RMF, ISO/IEC 23894, OWASP Top 10 LLM) and European regulatory standards, complemented by the analysis of the specific use case. The proposed solution integrates a serverless architecture comprising AWS Amplify, Amazon Cognito, API Gateway, AWS Step Functions, AWS Lambda, Amazon Bedrock (with Anthropic's Claude model), Amazon S3, AWS KMS, Amazon Textract, Amazon GuardDuty, Amazon CloudWatch, and AWS CloudTrail, with privacy-by-design and privacy-by-default applied at each layer. The end-to-end data flow analysis identifies five critical exposure points and proposes proportionate controls for each one; the AWS shared responsibility model precisely delimits which aspects of security and compliance fall on the provider and which fall on the customer, making clear that contracting managed infrastructure does not amount to transferring the responsibility held as data controller.

Results show that regulatory compliance and technical design are inseparable, that choosing a trusted cloud provider does not exempt the law firm from its obligations under the GDPR, and that human oversight constitutes the most important control in the system (both from the perspective of the lawyer's professional responsibility and from that of AI governance). The work concludes with a three-phase maturity roadmap: functional testing with synthetic data, a controlled pilot with scoped real data, and iterative expansion until normalized operation.

Keywords: Artificial Intelligence, AWS, GDPR, AI Act, Privacy by Design, Cloud Architecture, Legal Sector, AI Governance, Amazon Bedrock, MLOps.

Índice

Índice de ilustraciones

1. Introducción

1.1. Contexto y motivación

La inteligencia artificial (IA) ha dejado de ser una tecnología experimental para convertirse en un componente operacional de gran importancia en sectores intensivos en información, como son el legal y el financiero. Los despachos de abogados, las gestoras de fondos y los departamentos de cumplimiento normativo gestionan volúmenes crecientes de documentación compleja (contratos, escrituras societarias, expedientes fiscales, informes de due diligence) cuyo análisis manual resulta costoso, lento y propenso a errores humanos. En este contexto, los sistemas de IA orientados al procesamiento documental ofrecen la posibilidad de automatizar o asistir tareas que anteriormente requerían horas de revisión especializada (Jurafsky & Martin, 2023; Premasiri et al., 2025).

Sin embargo, la experiencia acumulada en los últimos años revela una brecha persistente entre la fase de experimentación y el despliegue productivo real. Construir un prototipo funcional o *proof of concept* (PoC) que extrae información de un conjunto de documentos de muestra y genera un resumen es técnicamente accesible con las herramientas actuales. Convertirlo en un sistema que opera de forma continua, segura, trazable y conforme a la normativa aplicable es una tarea de naturaleza completamente distinta, que implica dimensiones técnicas, organizativas, legales y éticas que el prototipo no aborda (Sculley et al., 2015; Amershi et al., 2019).

Esta brecha tiene consecuencias especialmente graves en el ámbito legal y financiero, donde los datos tratados incluyen información societaria confidencial, datos fiscales de personas jurídicas y físicas, y documentación protegida por el deber de confidencialidad inherente a la abogacía. En este entorno, un fallo de seguridad, una violación de privacidad o una respuesta incorrecta del sistema puede generar responsabilidades legales directas, pérdida de confianza del cliente y daño reputacional irreversible. El caso Moffatt v. Air Canada (2024), en el que un tribunal concluyó que la empresa era responsable de la información errónea proporcionada por su chatbot, ilustra el alcance de la responsabilidad que puede derivarse de un sistema de IA mal gobernado.

A todo ello se añade un marco regulatorio en rápida evolución. El Reglamento General de Protección de Datos (Reglamento (UE) 2016/679, en adelante "RGPD") establece obligaciones estrictas sobre el tratamiento de datos personales, incluyendo principios de minimización, limitación de finalidad, integridad y confidencialidad, y responsabilidad proactiva. El Reglamento de Inteligencia Artificial (Reglamento UE 2024/1689, en adelante "AI Act") introduce requisitos específicos para sistemas de IA en función del nivel de riesgo que presentan. Y el modelo de responsabilidad compartida de los proveedores cloud, como Amazon Web Services (AWS), exige que las organizaciones asuman activamente la responsabilidad de la configuración, el acceso y los controles de seguridad sobre la infraestructura que utilizan, sin que el proveedor les exima de sus obligaciones como responsables del tratamiento (Amazon Web Services, 2023).

El verdadero problema no es construir un prototipo de IA que analice documentación legal o fiscal, sino diseñar un sistema que pueda desplegarse en producción de forma segura, auditable y jurídicamente sostenible.

1.2. Problema de investigación y preguntas derivadas

A partir de la motivación anterior, el presente trabajo plantea la siguiente pregunta de investigación principal: ¿Cómo debe diseñarse y gobernarse una arquitectura cloud en AWS para desplegar en producción un sistema de IA que procese documentación sensible del ámbito legal cumpliendo con el RGPD, el AI Act y los requisitos de seguridad aplicables?

De esta pregunta principal se derivan cuatro subpreguntas que estructuran el desarrollo del trabajo. En primer lugar, qué condiciones previas debe satisfacer una organización antes de operar IA con datos sensibles. En segundo lugar, cómo se reparte la responsabilidad entre la organización, el proveedor cloud y el proveedor del modelo de IA. En tercer lugar, qué controles técnicos y organizativos son imprescindibles para garantizar privacidad, seguridad y auditabilidad. Y, en cuarto lugar, cómo evaluar si una arquitectura está realmente preparada para producción.

1.3. Objetivos generales y específicos

El objetivo general del presente trabajo consiste en diseñar una arquitectura de referencia auditable para el despliegue productivo de un sistema de IA sobre documentación societaria y fiscal sensible en AWS, garantizando el cumplimiento normativo, la seguridad y el gobierno del modelo.

Los objetivos específicos que articulan el logro del objetivo general son los siguientes. El primero es identificar los requisitos normativos, de seguridad y organizativos aplicables al caso de uso. El segundo es analizar los riesgos de privacidad, seguridad e IA asociados al sistema propuesto. El tercero es diseñar una arquitectura técnica justificada en AWS con servicios concretos y criterios de selección explícitos. El cuarto es definir controles de seguridad y privacidad por diseño que cumplan con el RGPD. El quinto es articular un marco de gobierno del modelo y de operación en producción. Y el sexto es evaluar la propuesta frente a criterios verificables de preparación para producción.

1.4. Alcance y limitaciones

El trabajo cubre el diseño arquitectónico del sistema, el análisis normativo aplicable, el gobierno del dato y del modelo, los controles de seguridad y privacidad, y el proceso de transición de PoC a producción. El caso de uso se acota a un despacho de abogados ficticio que procesa documentación societaria y fiscal de empresas analizadas en el marco de procesos de due diligence o asesoramiento.

1.5. Metodología

El trabajo adopta un enfoque de diseño arquitectónico basado en la revisión sistemática de literatura académica, marcos de referencia industriales y estándares normativos, complementado con el análisis del caso de uso concreto. La estructura metodológica sigue una lógica iterativa: análisis de requisitos, revisión del marco normativo y teórico, diseño de la arquitectura, análisis de riesgos y controles, y evaluación del diseño propuesto frente a criterios verificables. Esta aproximación es coherente con metodologías ágiles adaptadas a trabajos de carácter conceptual y técnico (Amershi et al., 2019).

2. Caso de uso, actores y requisitos del sistema

2.1. Descripción del caso de uso

El sistema objeto de este trabajo es una herramienta de asistencia jurídico-fiscal diseñada para un despacho de abogados especializado en operaciones societarias y asesoramiento fiscal. El caso de uso principal consiste en la revisión asistida con IA de la documentación societaria y fiscal de una empresa objetivo en el marco de procesos de due diligence, auditoría interna o asesoramiento continuado. En cada ejecución, el sistema procesa un conjunto de documentos relativos a la sociedad analizada (escrituras de constitución, modificaciones estatutarias, actas de órganos de gobierno, declaraciones fiscales, certificados de titularidad real e información de los socios y administradores) y asiste al abogado en la identificación de aspectos fiscales relevantes, como posibles contingencias, discrepancias entre documentos, estructuras accionariales complejas y otros elementos de interés para el análisis.

El sistema actúa como herramienta de asistencia, no como sistema de decisión autónoma. Todas las conclusiones generadas permanecen bajo la supervisión y responsabilidad del abogado que realiza el análisis. La inteligencia artificial no sustituye el criterio jurídico; lo complementa y acelera. Esta distinción entre asistencia y autonomía es determinante tanto para la clasificación del sistema bajo el AI Act como para la definición de los controles de gobernanza y supervisión.

2.2. Actores, roles y reparto de responsabilidades

La identificación precisa de los actores que intervienen en el sistema es el punto de partida para el análisis de accesos, trazabilidad y responsabilidad. La tabla siguiente resume los actores principales y su función:

Actor / Rol	Función en el sistema	Nivel de acceso
Abogado usuario	Carga documentos, lanza el análisis y revisa las salidas. Responsable del criterio jurídico.	Acceso a sus expedientes asignados
Administrador del sistema	Gestiona usuarios, permisos y configuraciones técnicas. No accede al contenido documental.	Administración de la plataforma
Responsable de seguridad y cumplimiento	Supervisa controles, revisa alertas y gestiona incidentes. Aprueba cambios de riesgo.	Logs y métricas de seguridad
AWS (proveedor cloud)	Encargado del tratamiento en cuanto a infraestructura. Opera bajo modelo de responsabilidad compartida.	Infraestructura subyacente
Amazon Bedrock / Anthropic (proveedor de modelos IA)	Proporciona el modelo de lenguaje. No usa inputs ni outputs del cliente para entrenamiento.	API de inferencia controlada

Tabla 1 Servicios, Funcionalidad y Nivel de acceso

La distinción entre la función del administrador (que no accede al contenido de los expedientes) y la del abogado usuario es una decisión de diseño deliberada que implementa el principio de separación de roles y reduce la superficie de exposición de información confidencial. El

modelo de responsabilidad compartida de AWS, que se analiza en detalle en el apartado 3.5, define el perímetro de responsabilidad del proveedor cloud respecto del despacho como responsable del tratamiento (Amazon Web Services, 2023).

2.3. Requisitos funcionales

Los requisitos funcionales del sistema definen qué debe hacer en términos de comportamiento observable. La tabla siguiente recoge los requisitos principales identificados a partir del análisis del caso de uso:

Requisito funcional	Descripción
Carga de documentos	El sistema acepta PDFs nativos, imágenes escaneadas y documentos de texto de las escrituras y expedientes fiscales.
Extracción de contenido textual	Se extrae el texto de los formatos mencionados, incluyendo documentos escaneados mediante OCR (Amazon Textract).
Análisis fiscal asistido	El sistema ejecuta un flujo estructurado de pasos de análisis fiscal, coordinado por AWS Step Functions.
Generación de informes	El sistema genera salidas estructuradas con conclusiones del análisis y referencias a los fragmentos documentales fuente.

Requisito funcional	Descripción
Registro de actividad	Toda acción relevante queda registrada: quién accedió, cuándo, qué operación realizó y qué respuesta generó el modelo.
Administración de usuarios y expedientes	El sistema permite gestionar usuarios, asignar permisos por expediente y controlar el ciclo de vida de los documentos.

Tabla 2 Requisitos Funcionales

2.4. Requisitos no funcionales

Los requisitos no funcionales del sistema son los que condicionan más directamente las decisiones arquitectónicas y tienen mayor impacto sobre el cumplimiento normativo. Se estructuran en las siguientes categorías:

Seguridad: el sistema debe implementar cifrado de datos en tránsito y en reposo, autenticación multifactor (MFA) para todos los usuarios, control de acceso granular basado en roles, y monitorización continua de amenazas y comportamientos anómalos.

Privacidad: el sistema debe aplicar minimización de datos en todas las etapas del procesamiento, limitación de finalidad que impida reutilizar los datos para otras finalidades no previstas, y mecanismos de borrado controlado y verificable de los expedientes.

Auditabilidad: toda acción relevante (accesos, cargas de documentos, invocaciones al modelo, respuestas generadas) debe quedar registrada con trazabilidad completa y conservarse de forma segura durante el período definido en la política de retención.

Rendimiento: los tiempos de respuesta del análisis deben ser adecuados al flujo de trabajo del despacho, con el análisis completo de un expediente disponible en un tiempo que no interrumpa la jornada de trabajo del abogado. En este caso, esto no es una prioridad para el sistema, la latencia puede ser relativamente alta,

Escalabilidad: el sistema debe ser capaz de procesar múltiples expedientes de forma simultánea sin degradación del servicio, adaptándose a picos de demanda sin reconfiguración manual.

Residencia del dato: todos los datos del sistema (documentos, embeddings, logs, resultados) deben almacenarse y procesarse exclusivamente en territorio de la Unión Europea.

Mantenibilidad: la arquitectura debe ser suficientemente modular para permitir la actualización de componentes individuales (incluyendo el modelo de IA) sin necesidad de rediseñar el sistema completo.

Explicabilidad razonable: las salidas del sistema deben incluir referencias explícitas a los fragmentos documentales que fundamentan cada conclusión, permitiendo al abogado verificar la fuente de cada afirmación del análisis.

2.5. Requisitos regulatorios y organizativos

Los requisitos regulatorios aplicables al sistema se derivan del RGPD, del AI Act y del marco de confidencialidad de la abogacía. Los más relevantes para el diseño son los siguientes:

- La existencia de base jurídica para el tratamiento de los datos personales que contienen los expedientes, típicamente el interés legítimo o la ejecución de un contrato de asesoramiento (art. 6 RGPD).
- La minimización de los datos enviados al modelo de IA, enviando solo el contenido estrictamente necesario para el análisis en cada paso.
- La limitación de finalidad, que impide reutilizar los datos del expediente para entrenar el modelo o para otras finalidades no declaradas.

- El control de acceso, que garantiza que solo los abogados autorizados acceden a cada expediente.
- La retención limitada, con eliminación de documentos conforme a la política de retención del despacho.
- Y la existencia de Acuerdos de Tratamiento de Datos (DPA) con AWS y con el proveedor del modelo, en su calidad de encargados del tratamiento según el artículo 28 del RGPD.
- Finalmente, el despacho debe mantener actualizado el Registro de Actividades de Tratamiento (RAT) con la descripción de este sistema.

3. Marco normativo y teórico

3.1. Fundamentos técnicos: NLP y sistemas documentales de IA

El sistema descrito se apoya en un conjunto de técnicas de procesamiento de lenguaje natural (NLP) y recuperación de información que conviene caracterizar brevemente para comprender los riesgos específicos asociados a cada capa. La primera etapa es la extracción de contenido: los documentos llegan en distintos formatos como pueden ser PDFs, imágenes escaneadas de escrituras notariales, formularios fiscales digitalizados y es necesario convertirlos a texto antes de cualquier procesamiento semántico. Para los documentos escaneados, se recurre al reconocimiento óptico de caracteres (OCR), que puede introducir errores en el texto extraído, especialmente en documentos de baja calidad, con sellos, firmas o formatos atípicos. La calidad del OCR condiciona directamente la fiabilidad de todo el análisis posterior (Jurafsky & Martin, 2023).

El modelo de lenguaje de gran escala (LLM) recibe como entrada la consulta del sistema y los documentos recuperados relevantes (más adelante se detalla cómo se determina la relevancia de estos documentos), y genera una respuesta estructurada. En el caso de un flujo de análisis fiscal, esta lógica se ejecuta en múltiples pasos encadenados, cada uno orientado a un aspecto específico: verificación de datos de constitución, análisis de la estructura accionarial, identificación de contingencias fiscales, entre otros. Esta arquitectura de procesamiento consiste en múltiples llamadas al modelo coordinadas por una orquestación con estados, introduciendo consideraciones adicionales de trazabilidad y control que se abordan en los capítulos 5 y 7 (NIST, 2024).

3.2. RGPD: principios aplicables al diseño del sistema

El Reglamento (UE) 2016/679 establece un conjunto de principios de protección de datos que deben traducirse en decisiones concretas de diseño cuando el sistema procesa datos personales. En el caso de un sistema de análisis fiscal societario, los datos tratados incluyen información de personas físicas (socios, administradores, representantes, titulares reales) que constituye

datos personales en el sentido del artículo 4.1 del RGPD, con todas las obligaciones correspondientes.

El principio de minimización de datos (art. 5.1.c) exige que solo se traten los datos estrictamente necesarios para la finalidad declarada, lo que se traduce en una política de extracción selectiva y en la limitación del contexto enviado al modelo en cada invocación. El principio de limitación de finalidad (art. 5.1.b) impide que los datos recopilados para el análisis fiscal se reutilicen para otras finalidades, incluyendo el entrenamiento del modelo de IA. Además, el principio de integridad y confidencialidad (art. 5.1.f) obliga a aplicar medidas técnicas y organizativas adecuadas para proteger los datos frente a tratamiento no autorizado, acceso ilícito o pérdida accidental, lo que fundamenta las decisiones de cifrado, control de acceso y monitorización continua de amenazas. Finalmente, el principio de responsabilidad proactiva (art. 5.2) exige que el responsable del tratamiento pueda demostrar el cumplimiento de todos los principios anteriores en cualquier momento, lo que crea una necesidad de documentación, registros de auditoría y evaluaciones de impacto continua.

El artículo 22 del RGPD regula las decisiones automatizadas que producen efectos jurídicos o significativamente similares sobre los interesados. En el caso de un sistema de asistencia (donde la decisión final siempre recae en el abogado) este precepto no resulta directamente aplicable. No obstante, conviene documentar explícitamente que el sistema no adopta decisiones autónomas y que el criterio jurídico es siempre responsabilidad del profesional, para evitar interpretaciones regulatorias problemáticas y garantizar alineación con las directrices del EDPB (2018).

3.3. Privacy by design & by default como metodología de diseño

El artículo 25 del RGPD consagra el principio de privacidad desde el diseño y por defecto (privacy by design y privacy by default), exigiendo que las medidas de protección de datos se integren en el sistema desde su concepción y que la configuración por defecto sea siempre la más protectora de la privacidad. Este principio tiene implicaciones directas en las decisiones arquitectónicas del presente trabajo y actúa como guía metodológica para el diseño de cada componente.

La privacidad por diseño implica adoptar medidas como: cifrado de datos en reposo y en tránsito como configuración por defecto y no como añadido opcional, arquitectura con separación de roles que impida físicamente el acceso no autorizado al contenido de los expedientes, reducción del contexto enviado al modelo al mínimo necesario para cada paso del análisis, y diseño de flujos de borrado que permitan eliminar datos de forma completa y verificable al vencimiento del período de retención.

La privacidad por defecto exige que, en ausencia de configuración explícita, el sistema adopte la opción más protectora de la privacidad. Esto se traduce, por ejemplo, en que el acceso a los expedientes esté deshabilitado por defecto para todos los usuarios y se habilite únicamente mediante asignación explícita de permisos por parte del administrador, o en que los logs de actividad registren únicamente metadatos y nunca el contenido de los documentos ni de las respuestas del modelo.

3.4. Reglamento de IA (UE) 2024/1689: enfoque de riesgo e implicaciones prácticas

El AI Act establece un marco de clasificación de sistemas de IA por nivel de riesgo, con requisitos escalonados en función de dicha clasificación. El reglamento entró en vigor el 1 de agosto de 2024, con aplicación progresiva: las obligaciones relativas a prácticas de IA prohibidas resultaron aplicables desde el 2 de febrero de 2025, las relativas a modelos de propósito general (GPAI) desde el 2 de agosto de 2025, y la plena aplicación de las disposiciones sobre sistemas de alto riesgo está prevista para el 2 de agosto de 2026 (Comisión Europea, 2024).

Es importante destacar que el sistema descrito en el presente trabajo (una herramienta de análisis fiscal bajo supervisión humana explícita y permanente) no encaja en las categorías de alto riesgo definidas en el Anexo III del AI Act, que incluyen sistemas utilizados en administración de justicia para tomar decisiones autónomas, o sistemas que afectan directamente a derechos fundamentales de personas físicas sin intervención humana significativa. La supervisión humana explícita, la naturaleza de asistencia del sistema y el perfil de usuarios (profesionales jurídicos que utilizan el sistema como apoyo a su criterio, no como

sustituto) son los elementos que determinan esta clasificación. Conviene, no obstante, documentar explícitamente esta valoración y mantenerla actualizada conforme evolucione el uso del sistema.

Las implicaciones prácticas del AI Act para el sistema son las siguientes:

- Necesidad de documentación sobre el propósito, las capacidades y las limitaciones del sistema
- Mecanismo explícito de supervisión humana que permita al abogado revisar, cuestionar y revertir cualquier salida del sistema (art. 14)
- Transparencia hacia los usuarios finales sobre el hecho de que interactúan con un sistema de IA
- Gobernanza interna que defina responsabilidades, procesos de aprobación de cambios y gestión de incidentes.

3.5. Responsabilidad compartida en la nube: el modelo AWS

Amazon Web Services opera bajo un modelo de responsabilidad compartida que delimita con precisión qué aspectos de la seguridad y el cumplimiento son responsabilidad de AWS y cuáles corresponden al cliente. Este modelo es fundamental para comprender las obligaciones reales que asume el despacho al utilizar AWS como infraestructura (Amazon Web Services, 2023).

AWS es responsable de la seguridad de la nube: protección física de los centros de datos, seguridad del hardware subyacente, del software de virtualización y de la red global. El cliente en este caso es responsable de la seguridad en la nube: configuración correcta de los servicios, gestión de identidades y accesos, cifrado de datos, configuración de la red virtual, y todos los controles aplicados sobre los datos y las aplicaciones desplegadas. En términos prácticos, esto significa que una configuración incorrecta de un bucket de Amazon S3, por ejemplo, dejarlo accesible al público, es responsabilidad del cliente, no de AWS.

En el plano del RGPD, AWS actúa como subencargado del tratamiento y ofrece el Acuerdo de Procesamiento de Datos (DPA) correspondiente, que incluye las garantías exigidas por el

artículo 28 del RGPD. Sin embargo, la responsabilidad de determinar la base jurídica del tratamiento, garantizar los derechos de los interesados y aplicar las medidas técnicas y organizativas adecuadas corresponde íntegramente al responsable del tratamiento. Entonces, se puede decir que elegir un proveedor cloud de confianza no equivale, en ningún caso, a cumplir con el RGPD.

3.6. Estándares y marcos de referencia

El presente trabajo se apoya en tres marcos de referencia complementarios para estructurar el análisis de riesgos y los controles propuestos. El AI Risk Management Framework del National Institute of Standards and Technology (NIST, 2023) propone un enfoque sistemático para identificar, evaluar y gestionar los riesgos de los sistemas de IA a lo largo de su ciclo de vida, articulado en cuatro funciones: gobernar, mapear, medir y gestionar. Este marco resulta especialmente útil para estructurar el gobierno del modelo descrito en el capítulo 7 del presente trabajo.

La norma ISO/IEC 23894:2023 proporciona guía específica de gestión de riesgos para sistemas de IA, compatible con la norma general ISO 31000 y aplicable tanto a riesgos técnicos —deriva del modelo, degradación de calidad, errores de extracción— como a riesgos de negocio y cumplimiento normativo (ISO, 2023). Su uso facilita la integración del análisis de riesgos de IA en los marcos de gestión de riesgos empresariales ya existentes en el despacho.

El OWASP Top 10 para aplicaciones con modelos de lenguaje de gran escala (OWASP, 2025) cataloga las diez categorías de riesgo más relevantes para sistemas basados en LLMs, incluyendo la inyección de prompts (manipulación del modelo mediante entradas diseñadas para alterar su comportamiento), la filtración de datos a través de las respuestas del modelo, la dependencia excesiva en el modelo sin supervisión humana adecuada, y los riesgos de la cadena de suministro de los modelos de IA. Este catálogo informa directamente los controles de seguridad descritos en los capítulos 5 y 6.

4. Análisis de riesgos, datos y gobierno

4.1. Inventario y clasificación de datos

El inventario de activos de datos es el punto de partida de cualquier análisis de privacidad y seguridad riguroso. Permite identificar qué información trata el sistema, cuál es su nivel de sensibilidad, cuál es la base jurídica que legitima su tratamiento y cuánto tiempo debe conservarse. La tabla siguiente recoge las categorías principales de datos que fluyen a través del sistema:

Categoría	Tipo de dato	Sensibilidad	Base jurídica
Documentación societaria	Escrituras de constitución, estatutos, modificaciones, actas de órganos	Alta (puede contener datos personales de socios y administradores)	Ejecución de contrato de asesoramiento (art. 6.1.b RGPD)
Información fiscal	Declaraciones fiscales, liquidaciones, certificados de la Agencia Tributaria	Alta (datos económicos protegidos y datos personales)	Ejecución de contrato (art. 6.1.b RGPD)
Datos de socios y administradores	Nombre, NIF/NIE, domicilio, porcentaje de participación, cargos	Alta (datos personales directamente identificables)	Interés legítimo o ejecución de contrato (art. 6.1.b/f RGPD)

Categoría	Tipo de dato	Sensibilidad	Base jurídica
Logs operacionales	Metadatos de acceso, identificadores de operación, marcas temporales, códigos de estado	Media (no contienen contenido documental)	Obligación legal / interés legítimo en seguridad (art. 6.1.c/f RGPD)
Resultados de análisis	Informes generados por el sistema con conclusiones del análisis fiscal	Alta (derivada del contenido documental fuente)	Ejecución de contrato (art. 6.1.b RGPD)

Tabla 3 Gobernanza del Dato

La categoría de mayor sensibilidad es la documentación societaria y la información fiscal, dado que concentra datos personales directamente identificables de personas físicas (socios, administradores, titulares reales) junto con información económica confidencial protegida por el deber de secreto profesional de la abogacía. Este nivel de sensibilidad justifica la aplicación de controles de máximo nivel en todas las etapas del ciclo de vida del dato, desde la carga hasta el borrado definitivo.

4.2. Flujo de datos extremo a extremo

El análisis del flujo de datos de extremo a extremo es fundamental para identificar los puntos de exposición del sistema y diseñar controles proporcionales en cada etapa. El recorrido del dato en el sistema sigue la secuencia siguiente.

El abogado accede al sistema a través del frontend desplegado en AWS Amplify, autenticándose mediante Amazon Cognito con autenticación multifactor obligatoria. Carga los documentos del expediente a través de la interfaz, que los transmite de forma cifrada (TLS 1.2 como versión mínima) hasta el backend. Los documentos se almacenan en Amazon S3, en un bucket con cifrado en reposo mediante AWS KMS y con acceso restringido mediante políticas IAM. Si el documento es una imagen escaneada o un PDF no textual, Amazon Textract extrae el contenido textual mediante OCR.

Cuando el abogado lanza el análisis, la petición atraviesa Amazon API Gateway y activa una AWS Step Function que orquesta el flujo completo del análisis. La Step Function coordina múltiples AWS Lambda, cada una responsable de un paso específico del análisis fiscal. Las Lambdas que requieren razonamiento sobre el contenido documental incluyen como contexto los documentos relevantes e invocan a Amazon Bedrock, que ejecuta un modelo de lenguaje (LLM) y devuelve la respuesta generada. Los resultados de cada paso se acumulan y, al finalizar el flujo, se genera el informe de análisis, que el abogado puede revisar en la interfaz. Toda la actividad queda registrada en Amazon CloudWatch y AWS CloudTrail.

Este análisis del flujo revela cinco puntos críticos de exposición que deben ser objeto de controles específicos: el canal de carga de documentos, el almacenamiento en S3 como concentrador de datos sensibles, el contexto enviado a Bedrock en cada invocación, los logs si registran más información de la necesaria, y el resultado del análisis si se almacena de forma no controlada.

4.3. Riesgos de privacidad y ciberseguridad

El análisis de riesgos identifica las principales amenazas para la privacidad y la seguridad del sistema. El acceso no autorizado a los expedientes es el riesgo de mayor impacto potencial: una configuración incorrecta de los permisos IAM, un compromiso de credenciales o una vulnerabilidad en la autenticación podría exponer documentación societaria y fiscal confidencial a personas no autorizadas. El control principal es el modelo de acceso basado en roles con mínimo privilegio y autenticación multifactor obligatoria.

La fuga de datos a través de los logs es un riesgo frecuentemente subestimado en sistemas de IA. Si los logs de CloudWatch o CloudTrail registran el contenido de los documentos procesados o de las respuestas del modelo, se crea una copia no controlada de información sensible que puede tener una retención y unas condiciones de acceso distintas a las del expediente original. El control es una política de logging estricta que registre únicamente metadatos como pueden ser identificadores, marcas temporales, códigos de estado, métricas de rendimiento y nunca el contenido.

El uso indebido del contexto enviado al modelo es un riesgo específico de las arquitecturas con IA: si el prompt incluye más información de la estrictamente necesaria para el paso del análisis en curso, se amplía innecesariamente la superficie de exposición ante el proveedor del modelo. Entonces, se debe establecer en el entorno una política de construcción de prompts que defina explícitamente qué información puede incluirse en el contexto y qué debe excluirse. La retención excesiva de expedientes aumenta también el riesgo de exposición en caso de incidente de seguridad: cuantos más datos se conservan durante más tiempo, mayor es el impacto potencial de una brecha. Por ello, también se configura una política de retención y borrado automático con verificación del borrado efectivo.

4.4. Riesgos específicos de IA

Los riesgos propios de los sistemas de IA presentan características distintas a los de la ciberseguridad clásica, ya que con frecuencia no generan un fallo obvio e inmediato, sino una degradación gradual o silenciosa de la calidad de los resultados, difícil de detectar sin monitorización específica.

Las alucinaciones son respuestas del modelo que parecen plausibles, pero son incorrectas o inventadas y constituyen el riesgo más relevante en el contexto legal. Un análisis fiscal que cite normas inexistentes, interprete incorrectamente un precepto o atribuya participaciones incorrectas puede inducir al abogado a error con consecuencias para el cliente. El caso Moffatt v. Air Canada (2024), en el que el tribunal concluyó que la empresa era responsable de información inexacta proporcionada por su chatbot, ilustra la responsabilidad que puede derivarse de la falta de control sobre las salidas del modelo. La arquitectura diseñadareduce este riesgo al anclar las respuestas en documentos reales, pero no lo elimina, especialmente

cuando el texto extraído por OCR contiene errores o cuando el fragmento recuperado no contiene la información relevante para la consulta.

La pérdida de trazabilidad se produce cuando no es posible determinar en qué documento se basó el modelo para generar una respuesta concreta. En un contexto legal, la capacidad de explicar y defender las conclusiones es esencial tanto para el abogado como para el cliente. El control es el diseño de las salidas del sistema para que incluyan siempre citas o referencias a los fragmentos fuente, permitiendo la verificación directa.

La dependencia del proveedor es un riesgo estructural inherente al uso de modelos comerciales de terceros: si el modelo cambia su comportamiento en una nueva versión, sus condiciones de uso o su disponibilidad en Amazon Bedrock, el sistema puede verse afectado de formas difíciles de anticipar. El diseño modular de la arquitectura con la lógica de negocio en las Step Functions y las Lambdas, y la invocación al modelo como un componente sustituible mitiga este riesgo al permitir cambiar el modelo con un impacto mínimo sobre el resto del sistema. Todo ello es facilitado por la plataforma, que con un cambio de una variable de identificación de modelo todo esto se puede actualizar.

5. Diseño de la arquitectura cloud en AWS

5.1. Principios de diseño

Las decisiones de diseño de la arquitectura propuesta se rigen por un conjunto de principios que actúan como criterios de evaluación de cada componente y que permiten justificar las elecciones realizadas frente a alternativas. Estos principios no son específicos de AWS, sino que reflejan las mejores prácticas de seguridad y privacidad por diseño aplicables a cualquier arquitectura cloud.

El principio de mínimo privilegio establece que cada componente del sistema (tanto usuario humano, función Lambda, rol de servicio) tiene acceso únicamente a los recursos que necesita para cumplir su función específica en el momento en que los necesita. El principio de segregación de entornos impone la separación física o lógica entre el entorno de desarrollo y el entorno de producción, impidiendo que los datos reales sean accesibles en entornos no productivos. El principio de cifrado por defecto exige que todos los datos estén cifrados en tránsito y en reposo sin necesidad de configuración adicional explícita en cada recurso. El principio de trazabilidad completa establece que toda acción relevante en el sistema queda registrada de forma inmutable. El principio de residencia del dato garantiza que todos los datos se almacenan y procesan en territorio de la UE. El principio de modularidad asegura que los componentes del sistema (incluido el modelo de IA) pueden actualizarse de forma independiente.

5.2. La plataforma AWS como infraestructura habilitadora

Amazon Web Services (AWS) es, a la fecha de redacción de este trabajo, una de las plataformas de computación en la nube con mayor amplitud de servicios gestionados especializados en inteligencia artificial, lo que la convierte en una opción especialmente pertinente para arquitecturas como la que se propone en este trabajo. En lugar de construir sobre componentes genéricos (máquinas virtuales, bases de datos de propósito general, librerías de código abierto sin soporte gestionado), AWS pone a disposición del desarrollador capas de abstracción de alta

madurez que reducen la complejidad operacional y permiten concentrar el esfuerzo en la lógica de negocio y los controles de cumplimiento.

En el ámbito del reconocimiento óptico de caracteres, Amazon Textract va más allá de la extracción de texto plano. El servicio analiza la estructura visual del documento e identifica no solo bloques de texto libre, sino también tablas con sus relaciones fila-columna, formularios con sus pares campo-valor, y jerarquías de secciones, lo que resulta especialmente valioso para documentación legal y fiscal con estructura compleja (Amazon Web Services, 2024a). Textract opera sobre una API asíncrona: el cliente envía el documento y recibe una notificación cuando el análisis ha concluido, lo que desacopla la fase de extracción de la lógica de aplicación y permite procesar documentos extensos sin bloquear al usuario. El servicio devuelve los resultados con coordenadas de posición para cada bloque detectado, lo que facilita la trazabilidad entre el texto extraído y el documento original.

Para el acceso a modelos de lenguaje de gran tamaño (LLM), Amazon Bedrock proporciona una API unificada que abstrae la complejidad de desplegar y escalar modelos de inferencia. A través de Bedrock, es posible invocar modelos de distintos proveedores (entre ellos Anthropic, Amazon, Meta y Mistral) bajo un único punto de entrada gestionado por AWS, sin necesidad de provisionar instancias de GPU, gestionar actualizaciones del entorno de ejecución ni mantener infraestructura de serving propia (Amazon Web Services, 2024b). La facturación se realiza por tokens consumidos, lo que alinea el coste con el uso real del sistema. Para el caso de uso de este trabajo, esta capacidad de invocar Claude de Anthropic a través de Bedrock en la región eu-south-2 (Madrid) permite garantizar la residencia del dato en territorio europeo y beneficiarse de las garantías contractuales de AWS en materia de privacidad, sin depender de una integración directa con la API del proveedor del modelo.

El resto de la plataforma complementa estos servicios con componentes que resuelven necesidades transversales de cualquier sistema productivo: AWS Lambda para ejecutar lógica de negocio sin gestión de servidores (*serverless*), con escalado automático y aislamiento entre ejecuciones; AWS Step Functions para orquestar flujos de trabajo complejos con gestión de estado, reintentos y bifurcaciones de error de forma declarativa; Amazon S3 para almacenamiento de objetos con una durabilidad de diseño de 99,999999999% y políticas de acceso granulares; Amazon Cognito para gestión de identidades de usuario con autenticación

multifactor y emisión de tokens JWT; y AWS KMS para gestión centralizada de claves de cifrado con registro de auditoría inmutable. La integración nativa entre todos estos servicios dentro de la misma plataforma reduce significativamente la superficie de integración que el equipo debe desarrollar y mantener, y facilita la aplicación coherente de controles de seguridad a lo largo de todo el sistema.

5.3 Arquitectura lógica

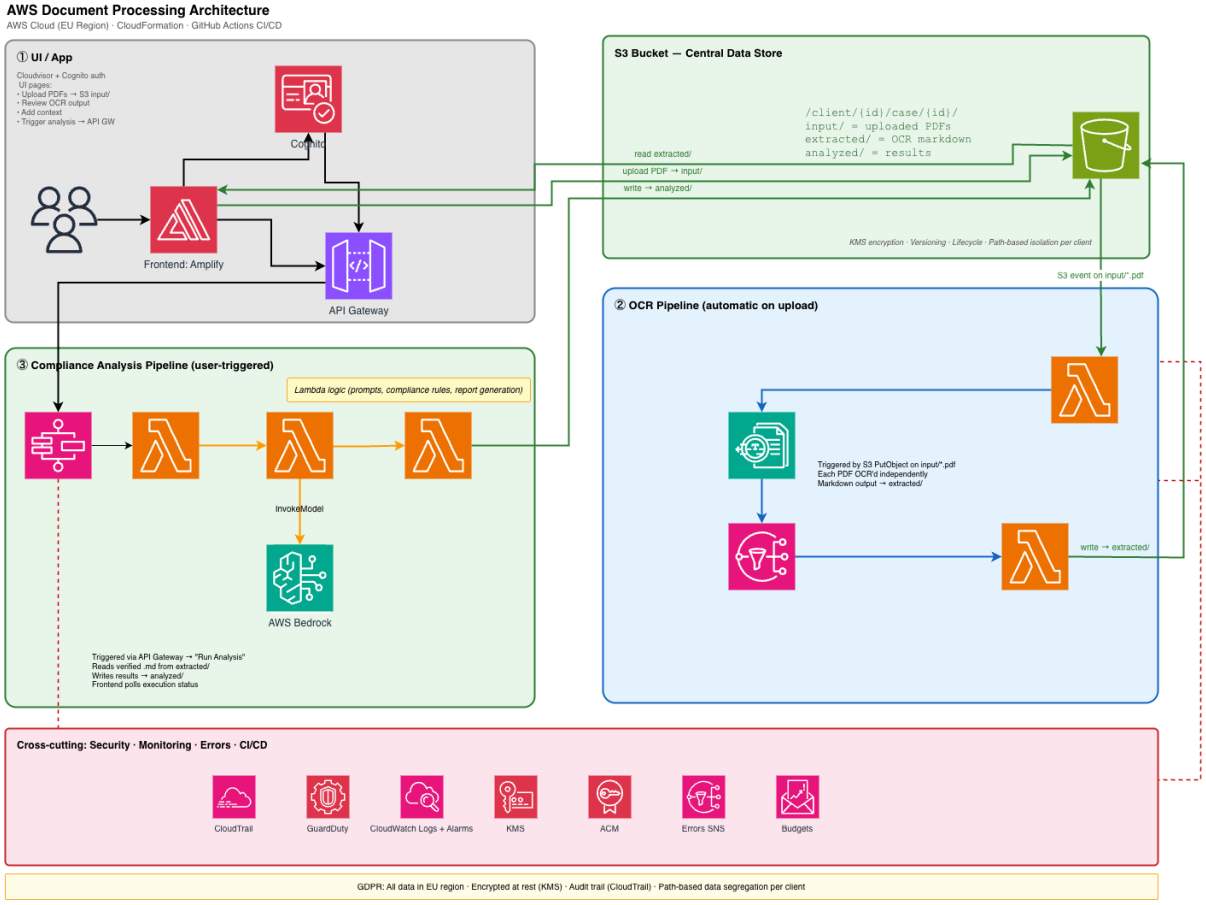

Ilustración 1 - Arquitectura en la Nube AWS

La arquitectura del sistema se organiza en tres bloques funcionales con responsabilidades bien delimitadas y fronteras de seguridad explícitas. Esta organización separa la lógica de extracción documental de la lógica de análisis de cumplimiento, lo que facilita el mantenimiento independiente de cada pipeline, permite que el OCR se ejecute de forma asíncrona sin bloquear al usuario, y hace posible auditar por separado cada tipo de operación.

El primer bloque es la capa de interfaz y autenticación, que gestiona la interacción del abogado con el sistema. El segundo bloque comprende dos pipelines de procesamiento: el pipeline de OCR, que se ejecuta automáticamente al cargar un documento, y el pipeline de análisis de cumplimiento, que el abogado dispara de forma explícita. El tercer bloque es la capa transversal de seguridad y monitorización, que atraviesa los dos pipelines anteriores y proporciona observabilidad, gestión de errores y control de costes.

Amazon S3 actúa como centro de datos del sistema, con una estructura de rutas jerárquica que implementa aislamiento lógico por cliente y por expediente: /client/{id}/case/{id}/input/ para los documentos originales subidos, /client/{id}/case/{id}/extracted/ para el texto extraído por OCR en formato Markdown, y /client/{id}/case/{id}/analyzed/ para los resultados del análisis generado por el pipeline de cumplimiento. Esta estructura de rutas es el mecanismo principal de aislamiento entre expedientes distintos y entre clientes del despacho, y se combina con políticas IAM y de bucket que restringen el acceso de cada componente a la ruta del expediente en curso.

5.4. Bloque I: capa de interfaz, autenticación y acceso

AWS Amplify proporciona el hosting del frontend de la aplicación web y gestiona el ciclo de vida de la interfaz de usuario. La interfaz permite al abogado realizar las operaciones principales del sistema: cargar los PDFs del expediente hacia la ruta input/ del bucket S3, revisar el resultado del OCR antes de lanzar el análisis, añadir contexto adicional si es necesario, disparar el análisis de cumplimiento e ir verificando cada paso. El frontend consulta periódicamente (polling) el estado de ejecución de la Step Function para informar al usuario del progreso del análisis sin necesidad de mantener una conexión permanente. La integración nativa de Amplify con Amazon Cognito simplifica la implementación de la autenticación, permitiendo que el frontend se beneficie directamente de los mecanismos de seguridad de Cognito sin lógica de autenticación propia en el cliente.

Amazon Cognito gestiona la identidad de los usuarios del sistema mediante un User Pool con autenticación multifactor (MFA) obligatoria para todas las cuentas. Cada abogado del despacho tiene una identidad propia en el pool, con acceso limitado a los expedientes asignados mediante la asignación a grupos con permisos diferenciados. Los tokens JWT emitidos por Cognito se

validan en Amazon API Gateway, garantizando que solo los usuarios autenticados y autorizados pueden acceder al backend.

Amazon API Gateway actúa como punto de entrada único y controlado al backend, configurado con el autorizador de Cognito para validar los tokens JWT en cada petición. API Gateway aplica limitación de tasa (rate limiting) para prevenir abusos, registra todas las peticiones en CloudWatch, y enruta las solicitudes hacia los componentes de backend correspondientes: la carga directa de documentos hacia S3 y las acciones relativas a la Step Function del pipeline de cumplimiento.

AWS Certificate Manager (ACM) gestiona los certificados TLS utilizados para cifrar las comunicaciones entre el navegador del abogado y los endpoints de Amplify y API Gateway. ACM automatiza la renovación de certificados antes de su expiración, eliminando el riesgo operacional asociado a la gestión manual de certificados y garantizando de forma continua el cifrado en tránsito sin intervención del equipo técnico.

5.5. Bloque II-A: pipeline de OCR

El pipeline de OCR se ejecuta de forma automática y asíncrona cada vez que el abogado sube un PDF al bucket S3, sin necesidad de intervención adicional. Este diseño event-driven desacopla la fase de extracción de texto de la fase de análisis, permitiendo que el OCR se procese en segundo plano mientras el abogado continúa trabajando.

El evento de creación de un objeto en la ruta input/*.pdf del bucket S3 dispara automáticamente la Lambda OCR Trigger, que recibe la notificación del evento S3, valida que el objeto es un PDF válido, e inicia el análisis asíncrono de **Amazon Textract**. Textract procesa el documento de forma asíncrona, que es especialmente relevante para documentos extensos como escrituras notariales de varias decenas de páginas y, al completar el análisis, notifica el resultado a través de Amazon SNS (Simple Notification Service). El tema SNS configurado para los callbacks de Textract dispara la Lambda Store .md, que recibe la notificación, recupera los resultados del análisis de Textract, convierte el contenido a formato Markdown estructurado, y escribe el archivo resultante en la ruta extracted/ del expediente en S3. Si se produce un error en cualquiera de los dos pasos del pipeline, tanto en la Lambda OCR Trigger como en la Lambda

Store .md, el flujo de error (representado en el diagrama con líneas discontinuas rojas) notifica el error a través del canal de gestión de errores del sistema. Cada PDF del expediente se procesa de forma independiente, lo que permite paralelizar la extracción de múltiples documentos y garantiza que un error en el procesamiento de un documento no bloquea el procesamiento del resto del expediente.

5.6. Bloque II-B: pipeline de análisis

El pipeline de análisis se activa cuando el abogado, tras revisar que el OCR de los documentos del expediente es correcto, pulsa la acción de ejecutar el análisis en la interfaz. La petición llega a través de API Gateway, que valida el token JWT y dispara la ejecución de la AWS Step Function que orquesta el flujo completo del análisis.

La Step Function coordina muchas Lambdas en secuencia, cada una con una responsabilidad bien delimitada. El pipeline de análisis se activa cuando el abogado, tras revisar que el OCR de los documentos del expediente, pulsa la acción de ejecutar el análisis en la interfaz. La petición llega a través de API Gateway, que valida el token JWT y dispara la ejecución de la AWS Step Function que orquesta el flujo completo. Este pipeline concentra la lógica de negocio propia del despacho: las reglas de cumplimiento, los criterios de análisis documental y la estructura del informe final son decisiones del equipo jurídico, no del proveedor de infraestructura. La Step Function coordina una secuencia de Lambdas que siguen una arquitectura de mini-servicios: cada función realiza un paso atómico y bien delimitado e invoca el modelo Claude de Anthropic a través de Amazon Bedrock cuando el paso requiere capacidad de inferencia sobre el texto extraído. En los puntos del flujo donde el resultado debe ser revisado antes de continuar, la Step Function entra en un estado de pausa mediante el mecanismo waitForTaskToken: la ejecución queda suspendida hasta que el abogado valida o rechaza el resultado desde la interfaz, garantizando así la supervisión humana en las decisiones de mayor impacto sin interrumpir la trazabilidad del estado del análisis.

Amazon Bedrock proporciona el acceso a modelos LLM mediante una API gestionada por AWS. Tras configurar el entorno adecuadamente, los inputs y outputs del cliente no se utilizan para entrenar modelos de AWS ni de terceros, el procesamiento se realiza en la región elegida, y los datos no son compartidos entre clientes (Amazon Web Services, 2024b). La invocación

al modelo se realiza especificando la versión exacta del modelo (model ID con versión), de forma que un cambio de versión requiere una decisión explícita y documentada.

5.7. Almacenamiento central: Amazon S3 y AWS KMS

Amazon S3 es el componente central de almacenamiento y el servicio que concentra los datos más sensibles del sistema. La estructura de rutas jerárquica /client/{id}/case/{id}/ implementa un aislamiento lógico robusto entre expedientes y clientes: las políticas de bucket garantizan que cada componente solo puede acceder a la ruta del expediente en curso, que el prefijo input/ solo es escribible por el usuario autenticado y los componentes del pipeline de OCR, que el prefijo extracted/ solo es escribible por la Lambda Store .md y legible por las Lambdas del pipeline de cumplimiento, y que el prefijo analyzed/ solo es escribible por las Lambdas de generación de informe. El acceso público está completamente bloqueado a nivel de cuenta y de bucket. El versionado de objetos está habilitado, lo que permite recuperar versiones anteriores en caso de modificación accidental y verificar la completitud del borrado al final del ciclo de vida del expediente. Las reglas de ciclo de vida de S3 automatizan la eliminación de objetos al vencimiento del período de retención definido en la política del despacho.

AWS KMS gestiona las claves de cifrado utilizadas para proteger los datos en reposo en S3. Para el cifrado del lado del servidor en S3 existen tres opciones principales. La primera es SSE-S3, en la que AWS genera, gestiona y rota las claves de forma completamente automática y opaca: el cliente no tiene visibilidad sobre qué claves protegen sus datos ni puede auditar cuándo o desde qué contexto se han utilizado. La segunda es SSE-KMS con claves gestionadas por AWS: las claves se crean automáticamente en KMS y el cliente puede ver los eventos de uso en CloudTrail, pero no puede modificar la política de la clave ni controlar qué roles o servicios están autorizados a usarla. La tercera es SSE-KMS con claves gestionadas por el cliente: el cliente crea la clave, define su política de acceso de forma granular, activa la rotación automática anual y obtiene un registro inmutable de cada operación de cifrado y descifrado en CloudTrail. Esta última opción es la elegida en la arquitectura propuesta, ya que es la única que permite restringir el uso de la clave a los roles de servicio concretos del sistema (y revocar ese acceso de forma inmediata si fuera necesario), lo que convierte el control de la clave en una capa de control de acceso independiente de las políticas de bucket. El cifrado SSE-KMS

con CMK se configura como política por defecto en todos los buckets, de forma que cualquier objeto almacenado queda cifrado automáticamente sin posibilidad de almacenamiento en claro.

5.8. Capa transversal: seguridad, observabilidad y errores

Amazon GuardDuty proporciona detección continua de amenazas mediante el análisis inteligente de los logs de CloudTrail, los registros de flujo de red y los logs de DNS. Detecta automáticamente comportamientos anómalos como pueden ser accesos desde ubicaciones inusuales o llamadas a la API fuera del patrón esperado y genera alertas que el responsable de seguridad debe revisar y gestionar según el procedimiento de incidentes.

AWS CloudTrail registra todas las llamadas a la API de AWS realizadas en la cuenta, proporcionando un registro inmutable y completo de toda la actividad de administración y configuración del sistema. Los registros de CloudTrail se conservan durante un mínimo de un año en un bucket de S3 de acceso restringido y son la fuente primaria de auditoría para demostrar el cumplimiento ante una inspección regulatoria.

Amazon CloudWatch recopila métricas y logs de todos los componentes del sistema: métricas de ejecución de las Lambdas, logs de la Step Function, logs de API Gateway y logs de aplicación generados por el código. Las alarmas de CloudWatch alertan al responsable técnico ante comportamientos que superen los umbrales predefinidos, errores frecuentes, latencias elevadas, fallos en el pipeline de OCR o en el de análisis.

Amazon SNS (Simple Notification Service) cumple dos funciones en el sistema. La primera es el mecanismo de callback asíncrono de Textract, que notifica a la Lambda Store .md cuando el análisis OCR de un documento ha completado. La segunda es la gestión centralizada de notificaciones de error: cuando se produce un fallo en cualquiera de los dos pipelines, el sistema publica un evento de error en un tema SNS configurado para enviar notificaciones por correo electrónico al responsable técnico, facilitando la detección temprana de problemas sin necesidad de monitorización activa continua.

AWS Budgets implementa el control de costes del sistema mediante alertas configuradas sobre el gasto mensual de la cuenta. Dado que el coste del sistema depende directamente del volumen

de análisis realizados (especialmente del número de páginas procesadas por Textract y del número de tokens consumidos en Bedrock), las alertas de Budgets permiten detectar incrementos de uso inesperados antes de que generen un impacto económico significativo, y proporcionan visibilidad sobre el coste real de operación del sistema a lo largo del tiempo.

5.9. Residencia del dato y elección de región

Todos los servicios del sistema se despliegan en la región AWS eu-south-2, correspondiente a España (Madrid). Esta elección garantiza que todos los datos del sistema se almacenan y procesan en territorio dentro de la Unión Europea, sin necesidad de aplicar los mecanismos de transferencia internacional del capítulo V del RGPD (Amazon Web Services, 2023). La disponibilidad en eu-south-2 de todos los servicios utilizados en la arquitectura hace posible un despliegue completamente regional sin transferencias de datos a otras regiones. Solo hay que remarcar que ciertos modelos de AWS Bedrock pueden ser ejecutados fuera de la región de España, pero no es un problema de cara a RGPD puesto que nunca abandonarían la Unión Europea. Y como indica la AEPD: las transferencias internacionales como flujos de datos desde España a destinatarios establecidos en países fuera del EEE no entran en el régimen de transferencias internacionales y no haría falta realizar SCCs o pedir autorización de la AEPD.

5.10. Gestión de identidades y control de accesos

La gestión de identidades combina Amazon Cognito para los usuarios finales y AWS IAM para los roles de servicio. Cada abogado tiene una cuenta propia en el User Pool de Cognito con MFA obligatoria. Los componentes del sistema utilizan roles IAM con permisos mínimos y explícitos: cada Lambda tiene un rol de ejecución propio que le permite acceder únicamente a las rutas S3, la clave KMS, el modelo Bedrock y el grupo de logs de CloudWatch estrictamente necesarios para su función. Ningún rol de servicio tiene permisos de administración de la cuenta ni acceso a expedientes ajenos al proceso en curso. Los accesos se revisan trimestralmente para detectar y eliminar permisos no utilizados.

5.11. Continuidad y resiliencia operacional

La arquitectura serverless basada en Lambda, Step Functions y S3 proporciona alta disponibilidad y escalabilidad automáticas sin gestión de servidores. En caso de fallo transitorio de una Lambda, la Step Function aplica la política de reintentos configurada para ese paso sin perder el estado del análisis. El pipeline de OCR procesa cada PDF de forma independiente, de forma que un error en el procesamiento de un documento no bloquea el resto del expediente. Los objetos S3 tienen una durabilidad de diseño de 99,999999999%. El versionado de objetos en los buckets de documentos permite recuperar versiones anteriores en caso de modificación accidental y verificar la completitud del borrado definitivo al final del ciclo de vida del expediente.

5.12. Modelo de responsabilidad compartida en AWS

Una de las premisas más relevantes del uso de servicios cloud en la gestión de datos sensibles es que contratar infraestructura gestionada no equivale a transferir la responsabilidad sobre la protección de datos ni sobre la seguridad del sistema. AWS articula esta realidad a través de su modelo de responsabilidad compartida (Shared Responsibility Model), que delimita de forma explícita qué aspectos de la seguridad y el cumplimiento corresponden al proveedor y cuáles corresponden al cliente (Amazon Web Services, 2023). La formulación canónica del modelo distingue entre la seguridad de la nube (responsabilidad de AWS) y la seguridad en la nube (responsabilidad del cliente). En la práctica, esta frontera no es idéntica para todos los servicios: varía en función del nivel de abstracción que cada servicio ofrece al cliente.

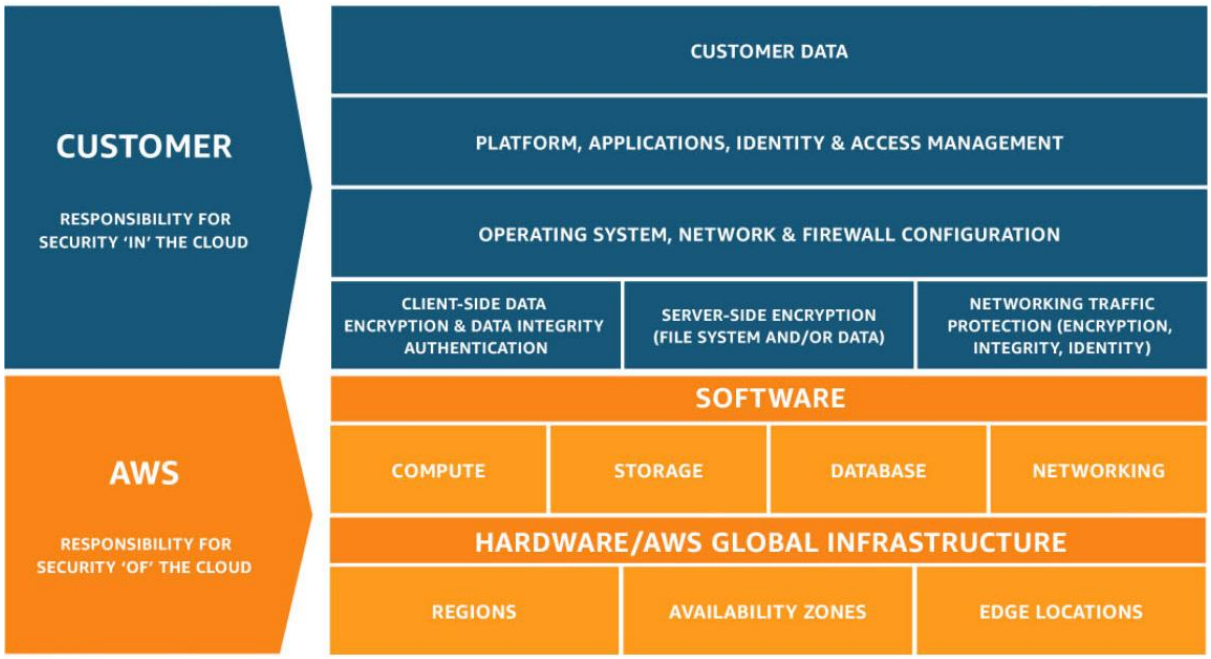

Ilustración 2 Modelo de Responsabilidad Compartida de AWS

En el nivel más profundo de abstracción, AWS asume la responsabilidad sobre la seguridad física de los centros de datos, la red global de interconexión de regiones y zonas de disponibilidad, el hardware de cómputo y almacenamiento, y el hipervisor o el entorno de ejecución gestionado de cada servicio. Estos componentes son inaccesibles para el cliente y nunca son objeto de configuración directa. AWS mantiene certificaciones de auditoría independiente sobre estas capas (entre ellas ISO 27001, SOC 2 Tipo II y el Esquema Nacional de Seguridad), que el cliente puede incorporar a su propio marco de cumplimiento sin necesidad de auditar la infraestructura física del proveedor (Amazon Web Services, 2023).

A partir de ahí, la frontera se desplaza hacia el cliente conforme se asciende en la pila de abstracción. Para los servicios utilizados en la arquitectura propuesta, el reparto de responsabilidades puede articularse de la siguiente forma:

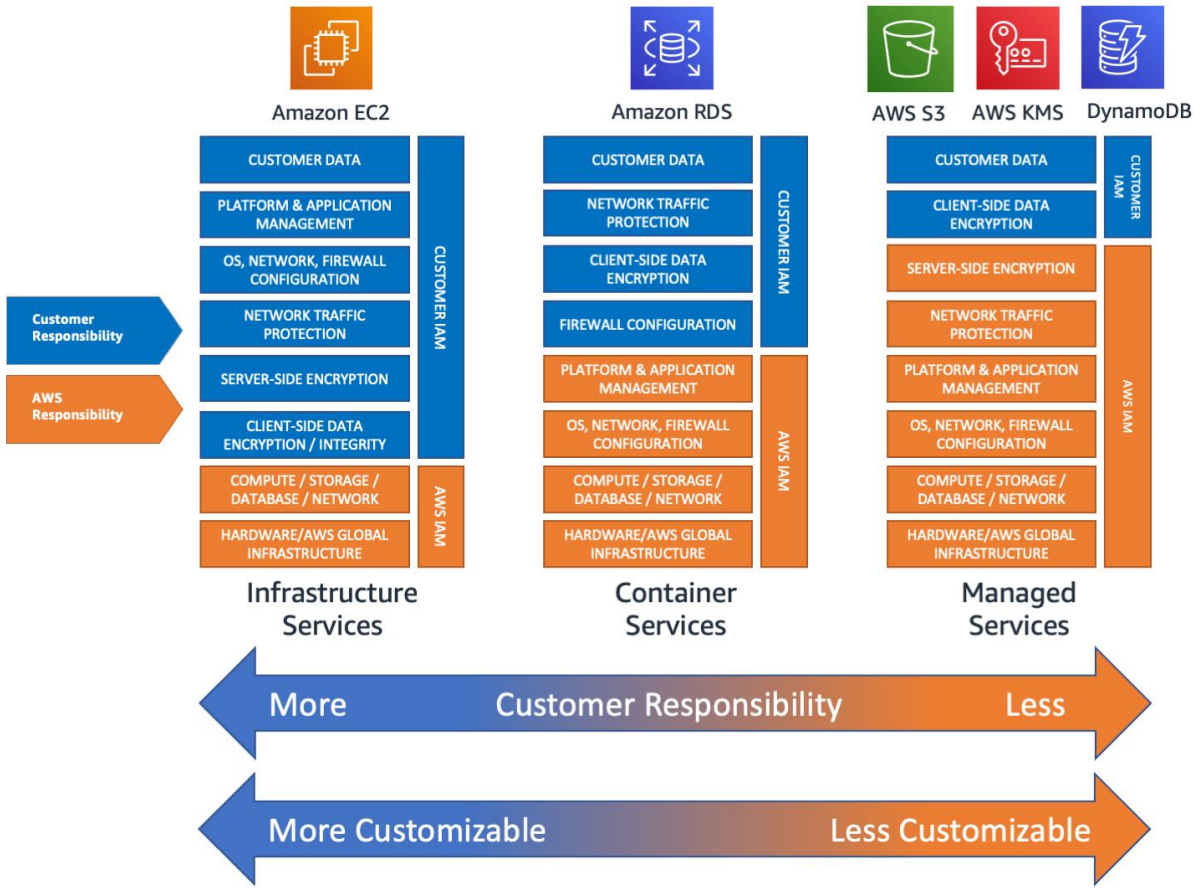

Ilustración 3 Responsabilidad de AWS según los servicios

Amazon S3: AWS garantiza la durabilidad y disponibilidad de los objetos almacenados, la seguridad física de las instalaciones y la correcta implementación del protocolo de cifrado. Sin embargo, la configuración del control de acceso (políticas de bucket, ACLs, bloqueo del acceso público, condiciones de las políticas IAM) es responsabilidad exclusiva del cliente. Un bucket S3 mal configurado, con acceso público habilitado o con una política demasiado permisiva, supone una vulnerabilidad que AWS no detectará ni corregirá de forma automática. Del mismo modo, aunque el sistema propuesto habilita el cifrado SSE-KMS por defecto en todos los buckets, AWS no impone esta configuración: es el equipo del despacho quien debe asegurarse de que ningún bucket del sistema puede almacenar objetos en claro.

AWS Lambda: AWS gestiona el entorno de ejecución de las funciones: el sistema operativo subyacente, el aislamiento entre ejecuciones de distintas funciones y las actualizaciones de seguridad del runtime. El cliente es responsable del código que se despliega en cada función,

de los permisos del rol de ejecución asociado, de la gestión segura de las variables de entorno, y de garantizar que los logs generados por la función no incluyen datos personales o sensibles de los expedientes. En la arquitectura propuesta, esto se traduce en que las Lambdas del pipeline de OCR y del pipeline de análisis deben implementar una lógica de logging que registre eventos de ejecución sin volcar el contenido de los documentos procesados.

Amazon Textract: AWS mantiene el modelo de OCR, la infraestructura de procesamiento y el canal seguro de comunicación con la API. El cliente es responsable de controlar qué documentos se envían a Textract y de gestionar correctamente los resultados devueltos: escribirlos en la ruta S3 correcta, con las políticas de acceso adecuadas, y no retenerlos en memoria o en logs más allá de lo estrictamente necesario. Textract, como servicio gestionado, no conserva los documentos procesados ni los resultados una vez entregada la respuesta (Amazon Web Services, 2024a), pero esta garantía contractual no exime al cliente de aplicar sus propios controles de acceso sobre los resultados almacenados en S3.

Amazon Bedrock: AWS gestiona el hosting del modelo, el escalado de la infraestructura de inferencia y las garantías contractuales sobre el uso de los datos: los inputs y outputs del cliente no se utilizan para entrenar modelos de AWS ni de terceros, y el procesamiento se realiza dentro de la región configurada (Amazon Web Services, 2024b). El cliente, sin embargo, es el único responsable del contenido que introduce en los prompts: debe garantizar que no se incluye más información personal de la estrictamente necesaria para el análisis, que los prompts están versionados y aprobados antes de su despliegue en producción, y que las respuestas del modelo son revisadas por un abogado antes de ser incorporadas a un informe. Bedrock no valida la adecuación jurídica de los prompts ni la exactitud de las respuestas generadas: esa responsabilidad recae íntegramente en el cliente.

AWS Step Functions: AWS garantiza la disponibilidad y la ejecución fiable del motor de orquestación. El cliente es responsable del diseño de la máquina de estados: las políticas de reintento configuradas para cada paso, la gestión de los estados de error, la definición de los timeouts y los permisos IAM que determinan qué recursos puede invocar cada estado de la Step Function. Un diseño deficiente de la lógica de reintentos puede provocar que, ante un error transitorio en Bedrock, el sistema reintente la llamada con el mismo prompt sobre datos

corruptos o incompletos, lo que generaría resultados incorrectos que podrían no detectarse sin una revisión humana adecuada.

Amazon Cognito: AWS gestiona la infraestructura del servicio de identidad y la correcta implementación de los estándares de autenticación (OAuth 2.0, OpenID Connect). El cliente es responsable de la configuración del User Pool: la política de contraseñas, la obligatoriedad del MFA, la duración de los tokens, la configuración de los clientes de aplicación y la asignación de usuarios a grupos con los permisos correspondientes. Cognito no impone por defecto el MFA obligatorio: es el administrador del sistema quien debe activarlo explícitamente para todas las cuentas.

AWS KMS: AWS garantiza la seguridad del servicio de gestión de claves y la correcta implementación del cifrado. El cliente es responsable de la política de cada clave: qué roles y qué servicios están autorizados a usar la clave para cifrar y descifrar, y qué acciones de administración de la clave están permitidas. Una política de clave demasiado permisiva (por ejemplo, que permita a cualquier rol de la cuenta usar la clave para descifrar) anularía el valor del cifrado SSE-KMS como control de acceso diferenciado.

En definitiva, la plataforma AWS proporciona una base sólida de seguridad en las capas de infraestructura y servicio gestionado, pero no elimina la necesidad de un diseño riguroso por parte del cliente. Para el caso de uso de este trabajo, esto significa que la conformidad con el RGPD y con el AI Act no se deriva de contratar AWS, sino del conjunto de decisiones de configuración, control de acceso, minimización de datos y supervisión humana que el equipo del despacho implementa y mantiene sobre esa infraestructura. El modelo de responsabilidad compartida debe entenderse, en consecuencia, no como una distribución de la carga de cumplimiento, sino como una delimitación de competencias técnicas: AWS garantiza que los cimientos son seguros; el cliente garantiza que el edificio está bien construido sobre ellos.

6. Seguridad y privacidad por diseño

6.1. Controles de seguridad por capas

La seguridad del sistema se organiza siguiendo un modelo de defensa en profundidad, con controles específicos y complementarios en cada capa de la arquitectura. Este enfoque garantiza que el compromiso de un control no elimina la protección del sistema, sino que es compensado por los controles de las capas adyacentes.

Capa de identidad: autenticación multifactor obligatoria para todos los usuarios, política de contraseñas robusta con renovación periódica, tokens JWT con tiempo de expiración corto, y revisión trimestral de accesos para detectar y eliminar cuentas o permisos no utilizados.

Capa de red: las Lambdas se despliegan con grupos de seguridad que restringen el tráfico entrante y saliente al mínimo necesario. API Gateway actúa como único punto de entrada público. Las comunicaciones entre servicios de AWS dentro de la misma región utilizan endpoints de VPC para evitar que el tráfico salga a Internet.

Capa de datos: cifrado en reposo con CMK en todos los buckets S3, acceso público completamente bloqueado, políticas de bucket con denegación explícita de accesos no autorizados, y gestión de ciclo de vida de objetos para eliminación automática al vencimiento de la retención.

Capa de aplicación: validación de entradas en todas las Lambdas para prevenir inyecciones, y política de construcción de prompts que minimiza el contexto enviado al modelo.

Capa de monitorización y respuesta: GuardDuty con detección continua de amenazas y alertas configuradas, CloudWatch Alarms para comportamientos anómalos en métricas operacionales, y CloudTrail para auditoría inmutable de administración.

6.2. Medidas de privacidad por diseño

Las medidas de privacidad por diseño se concretan en decisiones arquitectónicas específicas que implementan el artículo 25 del RGPD en cada capa del sistema. La minimización del contexto enviado al modelo es la medida más importante y la que mayor impacto tiene sobre la privacidad. La política de construcción de prompts define explícitamente qué información puede incluirse en el contexto y qué debe excluirse, y esta política se revisa en cada cambio significativo del flujo de análisis.

La limitación del logging de contenido impide que los logs de CloudWatch registren el contenido de los documentos o de las respuestas del modelo. Los logs registran únicamente metadatos, que son suficientes para la operación y la auditoría del sistema sin exponer información sensible. Esta decisión reduce significativamente el riesgo de fuga de información a través de los registros de actividad.

6.3. Controles organizativos

Los controles técnicos son necesarios pero insuficientes sin controles organizativos complementarios. La experiencia documentada en la literatura indica que la mayor parte de los incidentes de seguridad en entornos cloud se producen por error humano o por configuraciones incorrectas, no por vulnerabilidades en la infraestructura del proveedor (Sculley et al., 2015).

La política de uso del sistema define cómo deben utilizarlo los abogados: qué documentos pueden cargarse, qué usos son admisibles, cómo deben revisarse las salidas del modelo, y qué hacer en caso de detectar una respuesta incorrecta o sospechosa. Esta política se revisará anualmente o cuando se produzcan cambios significativos en el sistema. La formación periódica del personal garantiza que todos los abogados comprenden las capacidades y, especialmente, las limitaciones del sistema, y en particular que la supervisión humana es obligatoria y no opcional. Además, existe un proceso de gestión de incidentes que define cómo detectar, notificar, contener y resolver incidentes de seguridad o de calidad del sistema, con escalado al responsable de seguridad y, si procede, notificación a la Agencia Española de Protección de Datos en los plazos previstos por el RGPD.

6.4. Gestión de proveedores: AWS, proveedor del modelo y subencargados

El despacho actúa como responsable del tratamiento y debe garantizar, conforme al artículo 28 del RGPD, que sus proveedores, en calidad de encargados del tratamiento, ofrecen garantías suficientes de cumplimiento. AWS ofrece el Acuerdo de Procesamiento de Datos (DPA) estándar que incluye las garantías exigidas por el artículo 28 y las Cláusulas Contractuales Tipo aprobadas por la Comisión Europea, aunque dado que el sistema se despliega íntegramente en eu-south-2 y todos los datos permanecen en territorio español, la relevancia práctica de las transferencias internacionales es limitada en el diseño propuesto.

Amazon Bedrock actúa como intermediario hacia el LLM. Siguiendo la documentación pública de AWS, se configuró que los inputs y outputs del cliente no se usan para entrenar modelos de AWS ni de terceros, y los datos se procesan en la región configurada por el cliente (Amazon Web Services, 2024b). El despacho debe verificar explícitamente que el DPA de AWS cubre el uso de Bedrock como servicio de IA, que las condiciones de uso de Bedrock son compatibles con el RGPD y con el deber de confidencialidad de la abogacía, y que existen los mecanismos contractuales adecuados que regulen la cadena de tratamiento completa, incluyendo la posición de Anthropic como posible subencargado.

6.5. Evaluación de Impacto en la Protección de Datos (EIPD/DPIA)

El artículo 35 del RGPD exige realizar una Evaluación de Impacto en la Protección de Datos (EIPD, o DPIA por sus siglas en inglés) cuando el tratamiento, por su naturaleza, alcance, contexto o fines, entraña un alto riesgo para los derechos y libertades de las personas físicas. El sistema descrito procesa datos personales de socios y administradores de sociedades mediante técnicas de procesamiento automatizado avanzado, incluyendo modelos de IA, lo que sitúa el tratamiento en un área donde puede ser aconsejable realizar una EIPD con carácter preventivo. Sin embargo, la supervisión humana explícita y el carácter interno del sistema reducen la probabilidad de que sea obligatoria en sentido según los criterios establecidos.

En cualquier caso, la realización de una EIPD voluntaria reforzaría la posición del despacho ante una eventual inspección regulatoria, demuestra el compromiso con la responsabilidad proactiva exigida por el artículo 5.2 del RGPD, y permite identificar riesgos residuales que puedan requerir medidas adicionales. Los elementos que debe documentar una EIPD para este sistema incluyen: descripción sistemática del tratamiento y sus finalidades, evaluación de necesidad y proporcionalidad respecto a los fines, identificación y evaluación de los riesgos para los derechos y libertades de los interesados, y descripción de las medidas técnicas y organizativas previstas para mitigar dichos riesgos.

7. Gobierno de la IA y operación en producción

7.1. Criterios de selección del modelo

La elección del modelo de IA no es únicamente una decisión técnica basada en el rendimiento en benchmarks, sino también una decisión de gobernanza y cumplimiento con implicaciones sobre la privacidad, el control y la dependencia del proveedor. Los criterios relevantes para el caso de uso son los siguientes.

La capacidad técnica del modelo para comprender documentación jurídico-fiscal en español (como escrituras notariales, declaraciones del Impuesto de Sociedades, estatutos sociales) y para razonar sobre ella de forma coherente y fundamentada es el criterio técnico principal. Los modelos de Anthropic (Claude), accesibles a través de Amazon Bedrock, han demostrado un rendimiento superior en tareas de comprensión y síntesis de documentos legales complejos en español, lo que justifica su selección como modelo principal. La privacidad del tratamiento exige que el modelo no use los documentos del despacho para entrenamiento propio ni para mejorar versiones futuras del modelo. La combinación de Amazon Bedrock y los modelos de Anthropic ofrece esta garantía según la documentación pública disponible (Amazon Web Services, 2024b).

7.2. Protección del contenido enviado al modelo

La protección del contenido enviado al modelo es uno de los aspectos más críticos del sistema desde la perspectiva de la privacidad y del secreto profesional de la abogacía. Los documentos procesados contienen información confidencial de los clientes del despacho, cuya divulgación no autorizada vulneraría tanto el RGPD como las normas deontológicas aplicables a la abogacía.

Amazon Bedrock ofrece garantías contractuales relevantes: según su documentación, los inputs y outputs no se usan para entrenar modelos de AWS ni de terceros, el procesamiento se realiza en la región configurada por el cliente, y los datos no se comparten entre clientes ni se retienen más allá del tiempo necesario para procesar la solicitud (Amazon Web Services, 2024b). No

obstante, estas garantías contractuales son complementarias a los controles técnicos que el despacho debe aplicar, y no los sustituyen.

Los controles de diseño que protegen el contenido enviado al modelo son los siguientes. En primer lugar, la minimización del contexto: el sistema envía únicamente los fragmentos documentales relevantes para el paso del análisis en curso, no el expediente completo, con esto también ahorramos en el consumo de tokens. En segundo lugar, la ausencia de información de identificación directa en el prompt cuando no sea estrictamente necesaria para el análisis: si el paso del análisis puede realizarse sin incluir el nombre de la sociedad o el NIF de los socios, esa información se incluye.

7.3. Versionado de modelos, prompts y gestión del cambio

El versionado es una funcionalidad fundamental para la trazabilidad y la reproducibilidad del sistema. Cualquier cambio en el modelo, en los prompts o en el flujo de análisis puede alterar significativamente las salidas del sistema, lo que en un contexto legal tiene implicaciones directas sobre la consistencia y la calidad del asesoramiento prestado. Un abogado que obtiene resultados muy distintos de dos análisis del mismo expediente en momentos diferentes, sin que el expediente haya cambiado, no puede confiar en el sistema ni defender sus conclusiones ante un cliente o un tribunal.

El sistema implementa versionado en tres niveles. Los prompts de cada paso del análisis se almacenan en un repositorio de control de versiones con historia de cambios, autoría documentada y proceso de aprobación antes de su despliegue en producción. El modelo de Bedrock se invoca especificando la versión exacta del modelo (model ID con versión), de forma que un cambio de versión del modelo requiere una decisión explícita y documentada, no se produce de forma automática al actualizar el modelo en el proveedor. El flujo de la Step Function se versiona como código de infraestructura, con despliegues controlados mediante un proceso de revisión y aprobación que incluye la evaluación del impacto sobre la calidad de las salidas.

7.4. Supervisión humana y responsabilidad operativa

La supervisión humana es el control más importante del sistema y el que lo sitúa fuera de las categorías de alto riesgo del AI Act. El abogado que utiliza el sistema es responsable de verificar que los documentos cargados son los correctos y están completos, de revisar las salidas del análisis contrastando las conclusiones con los fragmentos documentales referenciados, de detectar y reportar respuestas incorrectas o sospechosas a través del procedimiento de incidentes, y de tomar la decisión jurídica final, que nunca puede delegarse al sistema.

Para que la supervisión humana sea efectiva y no meramente nominal, el sistema proporciona las herramientas necesarias para una revisión informada: cada conclusión del análisis va acompañada de referencias explícitas a los fragmentos documentales que la fundamentan, permitiendo al abogado verificar directamente la fuente. Por último, el sistema permite que el abogado puede corregir las salidas de los modelos en cada paso, para que en aquellos casos en los que los LLM no capturen toda la información, se puedan corregir los errores.

7.5. Itinerario de madurez: de PoC a producción por fases

El paso de un prototipo funcional a un sistema productivo requiere un proceso gradual que permita verificar el cumplimiento de los requisitos de seguridad, calidad y gobernanza antes de ampliar el alcance del sistema. La literatura sobre despliegue de IA en producción documenta que los proyectos que pasan directamente del prototipo al despliegue masivo sin fases intermedias tienen tasas de abandono y de incidente significativamente más altas que los que siguen un proceso de madurez gradual (Sculley et al., 2015; NIST, 2023).

El itinerario propuesto se estructura en tres fases que responden a un principio de expansión controlada del riesgo: se comienza probando el sistema en su totalidad sin datos reales, se valida después el comportamiento ante un caso de uso concreto con datos auténticos, y se culmina con la incorporación iterativa de nuevos procesos hasta alcanzar la operación normalizada.

La primera fase consiste en una prueba funcional integral con datos sintéticos o debidamente anonimizados. Su objetivo es verificar que el sistema funciona de extremo a extremo en todas sus variantes (ingesta, extracción, indexación, generación de respuestas, logging y trazabilidad) antes de que ningún dato real entre en contacto con él. El entorno debe estar completamente

aislado, sin conectividad con sistemas productivos ni credenciales sobre datos reales. Los criterios de paso a la siguiente fase incluyen la ausencia de errores críticos en el flujo principal, la cobertura de las variantes de uso previstas y la verificación de que los mecanismos de logging y rollback funcionan correctamente. La segunda fase introduce datos reales acotados a un único caso de uso, con revisión sistemática de todas las salidas por un profesional designado, cifrado activo, control de acceso basado en roles, logging operacional y DPA firmado con AWS. La tercera fase tiene carácter iterativo: a partir de ese caso validado, se incorporan progresivamente nuevas tipologías y volúmenes en ciclos sucesivos con evaluación intermedia, hasta que la supervisión pueda evolucionar hacia una revisión por muestreo, la gobernanza del modelo sea estable y los ciclos de revisión periódica formen parte de la rutina operativa de la organización.

8. Discusión

8.1. Principales hallazgos

El desarrollo del presente trabajo pone de manifiesto varios hallazgos de relevancia para cualquier organización que contemple el despliegue de sistemas de IA sobre documentación sensible en entornos cloud. El primero, y quizás el más importante, es que el cumplimiento normativo y el diseño arquitectónico no pueden separarse. Las decisiones técnicas, qué servicio gestiona el cifrado, cómo se construyen los prompts, dónde se almacenan los logs, qué información se envía al modelo… Tienen implicaciones directas y medibles sobre el cumplimiento del RGPD y del AI Act. Una organización que deja el análisis normativo para el final del proyecto, como validación de una arquitectura ya decidida, se enfrenta inevitablemente a revisiones costosas o, en el peor caso, a un sistema que no puede desplegarse legalmente.

El segundo hallazgo es que elegir un proveedor cloud de confianza no equivale a cumplir. El modelo de responsabilidad compartida de AWS deja en manos del cliente la configuración, el acceso, el cifrado y la gobernanza. La mayor parte de los incidentes de seguridad en entornos cloud no se producen por fallos del proveedor, sino por configuraciones incorrectas o insuficientes del cliente. Esto implica que la capacidad técnica del equipo que configura y opera el sistema es tan importante como la elección del proveedor.

El tercer hallazgo es que la gobernanza es tan importante como el modelo. Un sistema con un modelo de IA de alto rendimiento, pero sin versionado de prompts, sin procedimiento de incidentes, sin política de uso documentada y sin supervisión humana efectiva es un sistema que tarde o temprano fallará de formas que serán difíciles de detectar, explicar o corregir. La literatura sobre gestión de proyectos de IA en producción documenta consistentemente que los fallos más costosos no son fallos técnicos del modelo, sino fallos de gobernanza y operación (Sculley et al., 2015; Breck et al., 2017).

8.2. Aplicabilidad práctica

El marco propuesto es directamente aplicable a despachos de abogados, departamentos jurídicos internos de grandes empresas, gestoras de fondos que realizan due diligence sobre participadas, y consultoras que prestan servicios de asesoramiento fiscal o societario. En todos estos casos, el denominador común es la combinación de datos sensibles, obligaciones de confidencialidad, y necesidad de trazabilidad y auditabilidad. La arquitectura propuesta puede adaptarse a otros proveedores cloud (Azure o Google Cloud Platform) con ajustes en los servicios específicos sin cambios en los principios de diseño, los requisitos normativos ni el marco de gobernanza, lo que amplía su aplicabilidad más allá del ecosistema AWS.

9. Conclusiones

El presente trabajo ha respondido a la pregunta de cómo diseñar y gobernar una arquitectura cloud en AWS para el despliegue en producción de un sistema de IA que procese documentación legal y fiscal sensible, cumpliendo con el RGPD, el AI Act y los requisitos de seguridad aplicables.

La primera conclusión es que el despliegue de IA en producción sobre documentación sensible exige abordar simultáneamente cuatro dimensiones que no pueden tratarse de forma secuencial ni independiente: diseño técnico, cumplimiento normativo, seguridad y gobernanza. Ninguna de estas dimensiones puede compensar las deficiencias de las otras: un sistema técnicamente excelente sin gobernanza es operacionalmente frágil; un sistema con cumplimiento formal, pero sin controles técnicos es jurídicamente expuesto; un sistema seguro, pero sin supervisión humana efectiva es un riesgo para el cliente.

La segunda conclusión es que AWS, con los servicios descritos en el capítulo 5 de este trabajo (Amplify, Cognito, API Gateway, Step Functions, Lambda, Bedrock, S3, KMS, Textract, GuardDuty, CloudWatch y CloudTrail), proporciona una base técnica sólida para el cumplimiento del RGPD y del AI Act en el contexto del caso de uso planteado. Sin embargo, esta base solo es operativa si el cliente configura explícita y correctamente cada servicio siguiendo los principios de mínimo privilegio, cifrado por defecto, segregación de entornos y trazabilidad completa. El modelo de responsabilidad compartida no exime al despacho de ninguna de sus obligaciones como responsable del tratamiento.

La tercera conclusión es que la arquitectura basada en Step Functions y múltiples Lambdas que invocan Bedrock de forma coordinada es especialmente adecuada para el caso de uso planteado, porque permite organizar el análisis fiscal en pasos estructurados, modulares, trazables e independientemente actualizables. Esta modularidad es al mismo tiempo una virtud técnica (facilita el mantenimiento y la evolución del sistema) y una garantía de gobernanza que permite actualizar el modelo o los prompts de forma controlada sin afectar a toda la cadena.

La cuarta conclusión es que la supervisión humana no es un requisito burocrático que cumplir en papel para satisfacer al regulador, sino el control más importante del sistema desde la

perspectiva de la responsabilidad profesional del abogado. La herramienta de asistencia es tan buena como la calidad de la revisión humana que la acompaña. Diseñar el sistema para que esa revisión sea informada, eficiente y verificable es tan importante como diseñar la infraestructura técnica que lo sustenta.

En definitiva, poner IA en producción no es desplegar un modelo. Es diseñar un sistema gobernado, seguro, trazable y jurídicamente sostenible, capaz de operar de forma continua bajo la supervisión de los profesionales responsables de las decisiones que el sistema asiste.

10. Bibliografía

Amazon Web Services. (2023). Shared responsibility model. https://aws.amazon.com/compliance/shared-responsibility-model/

Amazon Web Services. (2024a). Amazon Textract developer guide. https://docs.aws.amazon.com/textract/latest/dg/what-is.html

Amazon Web Services. (2024b). Data protection in Amazon Bedrock. AWS Documentation. https://docs.aws.amazon.com/bedrock/latest/userguide/data-protection.html

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. En Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice (pp. 291-300). IEEE. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML test score: A rubric for ML production readiness and technical debt reduction. En 2017 IEEE International Conference on Big Data (pp. 1123-1132). IEEE. https://doi.org/10.1109/BigData.2017.8258038

Comisión Europea. (2024). AI Act: Entry into force and timeline of applicability. Digital Strategy. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Data Protection Board (EDPB). (2018). Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679 (revisadas en 2018). https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/automated-decision-making-and-profiling_en

Jurafsky, D., & Martin, J. H. (2023). Speech and language processing (3.a ed., borrador). Stanford University. https://web.stanford.edu/~jurafsky/slp3/

National Institute of Standards and Technology (NIST). (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

National Institute of Standards and Technology (NIST). (2024). Artificial intelligence: Generative AI. Considerations for responsible development and deployment (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

Open Web Application Security Project (OWASP). (2025). OWASP Top 10 for Large Language Model Applications (v2025). https://owasp.org/www-project-top-10-for-large-language-model-applications/

Organización Internacional de Normalización (ISO). (2023). ISO/IEC 23894:2023. Information technology. Artificial intelligence. Guidance on risk management. https://www.iso.org/obp/ui/es/#iso:std:iso-iec:23894:ed-1:v1:en

Premasiri, D., Ranasinghe, T., Mitkov, R., El-Haj, M., & Frommholz, I. (2025). Survey on legal information extraction: Current status and open challenges. Knowledge and Information Systems. https://doi.org/10.1007/s10115-025-02600-5

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Diario Oficial de la Unión Europea L 119, 1-88. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679

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 se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/232 (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea L 2024/1689. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J. F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. En Advances in Neural Information Processing Systems (Vol. 28, pp. 2503-2511). Curran Associates. https://proceedings.neurips.cc/paper_files/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

Tribunal de Resolución Civil de Columbia Británica (BCCRT). (2024). Moffatt v. Air Canada, 2024 BCCRT 149.

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Alberto García Fábregas, estudiante de [nombre del título] de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "DESARROLLO DE UN SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA LA GESTIÓN DE INFORMACIÓN EN EL SECTOR LEGAL Y FINANCIERO" declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Estudios multidisciplinares:** Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
4. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
5. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
6. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
7. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 22/04/2026

Firmado por GARCIA FABREGAS ALBERTO - ***2437** el día 22/04/2026 con un certificado emitido por AC FNMT Usuarios

Firma: _____