

**SISTEMA INTELIGENTE PARA LA AUTOMATIZACIÓN DE PROCESOS DOCUMENTALES  
USANDO IA, OCR Y RECUPERACIÓN SEMÁNTICA DE INFORMACIÓN**



**Profesor:**

**JULIÁN DANIEL TORRES VANEGAS**

**Alumnos:**

**CAMILO ANDRES CASCAVITA  
RODRIGUEZ YESSICA KATHERINE  
USECHE NIÑO**

**INGENIERÍA DE SISTEMAS**

**Bogotá D.C. Julio 2025**

## Tabla de Contenido

<b>Resumen Ejecutivo</b>	<b>4</b>
<b>Summary</b>	<b>4</b>
<b>Introducción</b>	<b>6</b>
<b>Antecedentes</b>	<b>7</b>
<b>Definición del problema</b>	<b>9</b>
<b>Objetivos</b>	<b>11</b>
<b>Justificación</b>	<b>12</b>
<b>Análisis de requerimientos</b>	<b>12</b>
<b>Marco de referencia</b>	<b>14</b>
<b>Diseño metodológico</b>	<b>16</b>
<b>Análisis de solución</b>	<b>19</b>
<b>Análisis de restricción</b>	<b>22</b>
<b>Alternativa de solución</b>	<b>25</b>
<b>Plan de implementación del sistema</b>	<b>28</b>
<b>Conclusiones</b>	<b>30</b>
<b>Referencias</b>	<b>31</b>
<b>Anexos</b>	<b>35</b>

## **Tabla de Figuras**

*Tabla 1. Análisis de costos, creación propia*

*Imagen 2. Login prototipo, creación propia*

*Imagen 3. Prototipo, creación propia*

*Imagen 4. Prototipo en ejecución, creación propia*

*Imagen 5. Pruebas de tiempos con llm, Creación propia*

## **Resumen Ejecutivo**

El presente proyecto tiene como objetivo desarrollar un sistema basado en inteligencia artificial (IA) para optimizar procesos documentales mediante la automatización de la extracción y análisis de información contenida en documentos. El sistema utilizará técnicas avanzadas de procesamiento de lenguaje natural y reconocimiento óptico de caracteres (OCR) para interpretar documentos en diversos formatos, como PDFs, archivos de Word e imágenes escaneadas, y responderá consultas específicas sobre su contenido de manera rápida y precisa.

El proyecto aborda una necesidad crítica en la gestión de información como lo es la pérdida de productividad causada por el tiempo dedicado a la búsqueda y además un procesamiento manual de documentos.

La solución propuesta optimizará la gestión documental en sectores como el legal, financiero, de recursos humanos y de salud, donde el manejo eficiente de grandes volúmenes de documentos es esencial para la toma de decisiones y el cumplimiento normativo.

## **Summary**

The objective of this project is to develop a system based on artificial intelligence (AI) to optimize document processes by automating the extraction and analysis of information contained in documents.

The system will use advanced natural language processing and optical character recognition (OCR) techniques to interpret documents in various formats, such as PDFs, Word files and scanned images, and will answer specific queries about their content quickly and accurately.

The project addresses a critical need in information management such as lost productivity caused by time-consuming search and manual document processing.

The proposed solution will optimize document management in sectors such as legal, finance, human resources and healthcare, where the efficient handling of large volumes of documents is essential for decision making and regulatory compliance.

## **Introducción**

En la era digital, la gestión eficiente de documentos es un desafío clave para las organizaciones debido al creciente volumen de información en formatos como PDF, Word e imágenes escaneadas. La búsqueda y procesamiento manual de documentos representa una pérdida significativa en términos de productividad, ya que los empleados pueden gastar hasta un 20% de su tiempo en estas tareas (McKinsey & Company, 2016).

En este contexto, la inteligencia artificial (IA) surge como una solución innovadora para automatizar procesos documentales, reducir errores y optimizar la recuperación de información. Este proyecto propone el desarrollo de un sistema basado en IA que emplea procesamiento de lenguaje natural (PLN), modelo de lenguaje de gran tamaño (LLM) y reconocimiento óptico de caracteres (OCR) para extraer, analizar y responder consultas sobre documentos de manera eficiente, minimizando la intervención humana.

El uso de IA en la gestión documental no solo mejora la eficiencia operativa, sino que también impulsa la transformación digital en las empresas, permitiendo a los profesionales enfocarse en actividades que no se encuentren enfocados en procesos documentales (Davenport & Ronanki, 2018).

Este proyecto no solo responde a una necesidad actual, sino que también sienta las bases para futuras investigaciones en IA aplicada a la gestión documental. Su implementación contribuirá a la reducción de costos operativos, la mejora en la precisión del manejo de la información y la optimización de procesos críticos en diversos sectores.

## **Antecedentes**

En los últimos años, la digitalización de documentos y la automatización de procesos han cobrado gran importancia en diversos sectores como el legal, financiero, educativo y de salud. La creciente necesidad de optimizar la gestión documental ha llevado a las organizaciones a implementar tecnologías emergentes, particularmente aquellas relacionadas con la inteligencia artificial (IA), el procesamiento de lenguaje natural (PLN) y el reconocimiento óptico de caracteres (OCR).

Según McKinsey & Company (2016), los empleados pueden dedicar hasta el 20 % de su jornada laboral a la búsqueda y gestión de información en documentos físicos o digitales no estructurados, lo que evidencia la urgencia de adoptar soluciones tecnológicas que aumenten la productividad organizacional. En esta misma línea, IBM (2018) destaca que las empresas pueden perder hasta un 30 % de su eficiencia operativa debido a la gestión manual de documentos.

El uso de OCR ha mejorado significativamente en la última década gracias a los avances en redes neuronales profundas. Herramientas como Tesseract OCR o Google Vision API han demostrado ser eficaces en la extracción de texto desde imágenes y documentos escaneados (Smith, 2007; Bharadiya, Thomas & Ahmed, 2023). No obstante, la simple digitalización no es suficiente; se requiere que los sistemas comprendan el contenido para responder preguntas y realizar búsquedas inteligentes.

En este contexto, los Modelos de Lenguaje de Gran Tamaño (LLM), como Llama o GPT, se han posicionado como herramientas clave para mejorar la interpretación semántica de los documentos (Brown et al., 2020). Su integración con bases de datos vectoriales como

FAISS o Qdrant permite una recuperación de información contextualizada y más precisa (Guu et al., 2020; Enholm et al., 2022).

Estudios recientes también han resaltado la tendencia creciente hacia la adopción de IA en procesos de gestión documental. Según Gartner (2022), para el año 2025, más del 50 % de las empresas implementarán soluciones basadas en inteligencia artificial para optimizar sus flujos documentales.

Estos antecedentes reflejan el contexto tecnológico y empresarial que da origen al presente proyecto, cuyo objetivo es diseñar un sistema inteligente que integre OCR, LLM y bases vectoriales para automatizar la extracción, análisis y consulta de información documental de forma eficiente y segura.

## **Definición del problema**

En diversos entornos académicos y empresariales, la gestión eficiente de documentos es un desafío recurrente debido al gran volumen de información que debe ser procesada, almacenada y consultada de manera rápida y precisa. Muchas veces, estos documentos se encuentran en formatos no estructurados (PDFs escaneados, imágenes, archivos de texto, entre otros), lo que dificulta su búsqueda y análisis.

A pesar de los avances tecnológicos, muchas industrias aún dependen de soluciones tradicionales de gestión documental, las cuales presentan limitaciones en la extracción eficiente de información y en la consulta automatizada de documentos. En sectores como el financiero, legal y de salud, donde la documentación es extensa y de vital importancia para la toma de decisiones, la búsqueda y análisis manuales representan un desafío, consumiendo tiempo y recursos significativos (Pure Storage, 2025).

El sistema propuesto busca optimizar estos procesos mediante inteligencia artificial, permitiendo una gestión más eficiente y precisa de documentos críticos en estos sectores.

Un informe de IBM (2018) revela que las organizaciones pueden perder hasta un 30% de su tiempo operativo debido a la ineficiencia en la gestión de documentos. Esta ineficiencia se traduce no solo en la búsqueda y recuperación de información, sino también en el almacenamiento, organización y archivo de documentos de forma manual. En empresas con grandes volúmenes de información, estos procesos resultan ser aún más desafiantes, pues la falta de un sistema adecuado para organizar, indexar y recuperar datos puede derivar en pérdidas importantes de información, reducción en la calidad de las

decisiones y un aumento en el riesgo de incumplir con normativas regulatorias, como las de protección de datos personales (Hernández et al., 2020)

En este contexto, la implementación de una capa de inteligencia artificial que integre modelos de OCR, bases vectoriales y modelos de lenguaje (LLMs) puede representar una solución innovadora para optimizar la consulta y análisis de documentos. Sin embargo, el desarrollo de una IA completa requiere múltiples etapas de evaluación y recursos, por lo que es necesario enfocarse en el diseño de un prototipo funcional que demuestre la viabilidad de esta tecnología.

El presente proyecto busca diseñar un aplicación que combine estas herramientas, permitiendo la extracción y recuperación eficiente de información en documentos mediante inteligencia artificial. Esto facilitará la interacción con grandes volúmenes de datos de manera automatizada y semántica, reduciendo los tiempos de consulta y mejorando la accesibilidad de la información.

## **Objetivos**

### **Objetivo General:**

Desarrollar un prototipo de sistema basado en inteligencia artificial para la optimización de procesos documentales, a través de la extracción, análisis y consulta eficiente de información contenida en documentos de diversos formatos.

### **Objetivos específicos:**

- Definir la arquitectura del sistema, estableciendo sus principales componentes, incluyendo el frontend, los modelos de OCR, bases vectoriales y modelos de lenguaje (LLM).
- Implementar la integración de modelos OCR para la conversión de documentos en distintos formatos (PDF, imágenes, archivos de Word) en datos estructurados, optimizando la precisión y velocidad del reconocimiento de texto.
- Diseñar una interfaz de usuario intuitiva y accesible, que facilite la consulta de documentos y la obtención de respuestas basadas en IA y/o LLM.

## **Justificación**

A medida que las empresas generan y almacenan grandes volúmenes de datos, el procesamiento manual se vuelve cada vez más complejo, esto genera un aumento significativo en los costos asociados a tiempo y recursos. Además, esta gestión manual conlleva un alto riesgo de errores humanos, lo que puede afectar negativamente la calidad de los procesos y la toma de decisiones. Según un estudio de McKinsey & Company (2016), los empleados destinan hasta el 20% de su jornada laboral a la búsqueda y gestión de información, lo que representa una considerable pérdida de productividad.

Ante esta problemática, la inteligencia artificial (IA) ofrece una solución innovadora, que va más allá de la simple automatización. El uso de técnicas avanzadas de procesamiento de lenguaje natural y reconocimiento óptico de caracteres reduce la intervención humana y permite mejorar la precisión de los resultados. Con un sistema de IA correctamente implementado, las organizaciones pueden no solo optimizar el tiempo destinado a la gestión documental, sino también reducir errores, mejorar la coherencia y congruencia de los datos para así garantizar una mayor fiabilidad en la toma de decisiones.

Un sistema de gestión documental basado en IA ayudaría en la mejora de la toma de decisiones dentro de la organización. Ya que se tendrían datos estructurados, organizados y actualizados en tiempo real, las empresas tienen acceso a una fuente confiable de información que les permite tomar decisiones más rápidas y precisas. Este acceso inmediato y eficiente a datos procesados y correctamente clasificados facilita la identificación de patrones y tendencias, lo que puede ofrecer ventajas competitivas al predecir cambios en el mercado o en el comportamiento de los clientes. Además, que a

través de estas herramientas se pueden realizar análisis de distintas tendencias, papel clave para la toma de decisiones. Según Davenport y Ronanki (2018), la automatización

permite que los profesionales se enfoquen en actividades estratégicas de mayor valor agregado, permitiendo mayor competitividad profesional.

Por otra parte, un informe de Gartner (2022), empresa estadounidense enfocada en la consultoría e investigación de las tecnologías de la información, proporcionando los estándares de vanguardia tecnológica; menciona que para 2025 más de la mitad de las empresas implementarán soluciones basadas en IA para la gestión documental, lo que resalta la importancia de este tipo de proyectos en la transformación digital de las organizaciones.

Entre los beneficios esperados del sistema propuesto se encuentran:

- **Reducción de costos operativos**, al minimizar y optimizar la intervención humana en la gestión documental.
- **Optimización del tiempo**, al permitir una búsqueda y extracción de información de una manera más eficiente y eficaz
- **Mejora en la toma de decisiones**, gracias a que los datos estructurados serían más precisos y se podrían gestionar en tiempo real
- **Incremento en la productividad**, al reducir el tiempo dedicado a tareas repetitivas y enfocarse en actividades estratégicas.
- **Transformación digital**, alineando a las organizaciones con las tendencias actuales de automatización e inteligencia artificial, posicionando la empresa en un entorno cada vez más cambiante

Este proyecto no solo tiene el potencial de mejorar la eficiencia operativa en distintos sectores, sino que también puede servir como base para futuras investigaciones y desarrollos en el ámbito de la IA aplicada a la gestión documental.

## **Marco de Referencia**

El marco teórico de este proyecto se basa en la revisión y análisis de investigaciones previas, teorías y normas relacionadas con la gestión documental, la inteligencia artificial (IA), Modelos de lenguajes grandes (LLM) y el reconocimiento óptico de caracteres (OCR).

Este marco proporciona información importante para el desarrollo del sistema propuesto, que busca optimizar la gestión de documentos mediante la automatización de la extracción y análisis de información. La gestión documental es un proceso crítico en la mayoría de organizaciones, especialmente en sectores como el legal, financiero, recursos humanos y de salud, donde se hace necesario el manejo eficiente de grandes volúmenes de documentos, todo esto en busca de tener mayor información a la hora de tomar decisiones (Pure Storage, 2025).

La inteligencia artificial (IA), se ha posicionado como un herramienta clave y fundamental en procesos de automatización demostrando ser importante para la interpretación y análisis de textos. Los modelos de lenguaje avanzados, como Llama 3.1, Phi-3 y Gemma 2 o 3, permiten la comprensión contextualizada de documentos, lo que en principio facilita la extracción de información relevante y la respuesta a consultas específicas (Bharadiya, Thomas, & Ahmed, 2023).

Estos modelos de gran escala han transformado la manera en cómo se recupera la información, teniendo como premisa respuestas coherentes y adaptadas al contexto, ya que estos modelos tiene la posibilidad de realizar tareas sin un entrenamiento previo lo que permite mejorar la eficiencia en el procesamiento documental (Brown et al. 2020).

La integración de bases de datos vectoriales, como Qdrant o FAISS, permite optimizar la recuperación de información basada en IA, permitiendo una búsqueda semántica más eficiente (Enholm et al., 2022). Su aplicación en la gestión documental mejora la búsqueda y clasificación de los documentos, todo esto a partir de las representaciones vectoriales, permite además búsquedas semánticas que resultan ser más precisas (Johnson et al. 2019)

La utilización de reconocimiento óptico de caracteres (OCR) es clave para la conversión de documentos. Estas herramientas han mejorado significativamente la precisión y velocidad de esta conversión, lo que permite una gestión más eficiente de documentos en formatos no estructurados (Bharadiya et al., 2023). Los avances en redes neuronales han permitido una significativa mejora en la precisión de los OCR y además una excelente conexión con sistemas de IA (Smith, 2007).

La implementación de soluciones basadas en IA no solo contribuye a la eficiencia operativa, sino que además permite a las distintas organizaciones ser más competitivas en un mercado global. De acuerdo con McKinsey & Company (2023), la automatización mediante el uso de herramientas enfocadas en IA ha permitido que las empresas reduzcan significativamente el tiempo dedicado a tareas administrativas, lo que evidencia su impacto en la optimización de procesos empresariales.

Este proyecto no solo aborda una problemática actual, sino que también proporciona las bases para investigaciones y desarrollos en IA que se encuentren dedicados a la gestión documental. Al combinar distintas técnicas, se crea un sistema escalable y adaptable que pueda no solo utilizarse en gestión documental sino que además pueda implementarse en diferentes contextos y sectores.

## **Diseño Metodológico**

Para el desarrollo del presente proyecto se adoptó un enfoque metodológico mixto, combinando elementos cualitativos y cuantitativos con un diseño descriptivo y aplicado. Esta elección permite abordar tanto los aspectos técnicos del sistema como la experiencia de usuario en contextos reales. La metodología empleada se fundamenta en el marco de Design Science Research (DSR), el cual ha sido ampliamente utilizado en proyectos de ingeniería de software e innovación tecnológica (Hevner, March, Park & Ram, 2004).

El enfoque DSR resulta adecuado porque permite iterar entre fases de diseño, implementación y evaluación, facilitando la construcción de artefactos tecnológicos útiles y funcionales. Esta metodología garantiza que el sistema propuesto no solo sea técnicamente viable, sino que también responda a necesidades reales en contextos organizacionales específicos.

### **Enfoque y fases del proyecto**

El proyecto se estructura en cinco fases principales:

- **Revisión bibliográfica y análisis normativo:** Se realizó una revisión sistemática de literatura académica y técnica sobre tecnologías como OCR, LLM y bases de datos vectoriales. Asimismo, se revisaron las normativas legales relevantes, como la Ley 1581 de 2012 y el Reglamento General de Protección de Datos (GDPR), lo que permitió definir los requisitos funcionales y no funcionales del sistema (European Union, 2016; Congreso de Colombia, 2012).
- **Diseño del sistema:** Se definió la arquitectura técnica del sistema, integrando un frontend desarrollado en Streamlit, un backend con módulos de OCR y modelos de

lenguaje, y una base de datos vectorial. Esta etapa incluyó la elaboración de diagramas UML y wireframes para facilitar la validación temprana con stakeholders (Treveil et al., 2020).

- **Implementación del prototipo:** Se construyó un prototipo funcional utilizando herramientas como Tesseract OCR, LangChain, Hugging Face y PyTorch. Se realizaron pruebas iniciales con documentos reales para ajustar el reconocimiento óptico y mejorar la precisión de los modelos (Smith, 2007; Bharadiya, Thomas & Ahmed, 2023).
- **Evaluación del sistema:** Se llevaron a cabo pruebas de precisión del OCR, medición de latencia del sistema y evaluación de usabilidad mediante la escala SUS. Además, se aplicaron encuestas y entrevistas semiestructuradas a usuarios con experiencia en gestión documental, siguiendo las recomendaciones metodológicas de Hernández, Fernández y Baptista (2020).
- **Iteración y validación final:** Con base en los resultados obtenidos, se realizaron ajustes en la interfaz y el backend. Se elaboró documentación técnica, incluyendo manual de usuario y guía para desarrolladores. Finalmente, se socializaron los resultados con actores clave para su retroalimentación.

### **Técnicas de recolección y análisis de datos**

Se emplearon técnicas de recolección tanto cuantitativas como cualitativas. A nivel cuantitativo, se midieron variables como la tasa de reconocimiento de texto (precisión del OCR), el tiempo de respuesta del sistema y el porcentaje de aciertos en respuestas generadas por el modelo. Estas métricas se analizaron con herramientas estadísticas en Python, usando librerías como Pandas y SciPy.

En el plano cualitativo, se analizaron las respuestas de los participantes utilizando codificación temática con el apoyo del software NVivo, permitiendo identificar patrones recurrentes en la experiencia de uso del sistema. Esta triangulación metodológica permitió validar no solo la eficacia técnica del prototipo, sino también su pertinencia práctica y aceptación por parte de los usuarios finales.

## **Alternativa de solución**

En el proceso de diseño del sistema inteligente para la gestión documental, se consideraron múltiples alternativas tecnológicas con el fin de identificar la solución más viable, eficiente y adaptable a los contextos empresariales reales. Las alternativas fueron evaluadas con base en criterios como escalabilidad, facilidad de integración, costos de implementación, precisión técnica y cumplimiento normativo.

### **1. Sistemas tradicionales de gestión documental**

Una de las primeras opciones evaluadas fueron los sistemas de gestión documental (DMS) tradicionales, como Alfresco, OpenText o SharePoint. Estas plataformas ofrecen almacenamiento, clasificación y búsqueda básica de documentos mediante metadatos o etiquetas. Sin embargo, presentan limitaciones al trabajar con documentos no estructurados (como imágenes o escaneos) y no incorporan inteligencia semántica para consultas avanzadas (Pérez & Vargas, 2021). Además, su personalización para tareas de extracción automatizada de datos suele requerir altos costos de configuración y mantenimiento.

### **2. Digitalización básica con OCR sin IA**

Otra alternativa considerada fue el uso exclusivo de motores OCR para convertir documentos físicos o escaneados en texto editable. Herramientas como ABBYY FineReader o Tesseract permiten digitalizar documentos de forma precisa (Smith, 2007). No obstante, estas soluciones no ofrecen funcionalidades de comprensión del texto ni recuperación semántica, lo que impide realizar búsquedas contextuales o responder

preguntas complejas sobre el contenido de los documentos (Patel, Sharma & Gupta, 2021).

### **3. Plataformas basadas en RPA (Automatización Robótica de Procesos)**

También se analizaron soluciones basadas en RPA como UiPath o Automation Anywhere, que permiten automatizar tareas repetitivas relacionadas con la gestión documental (por ejemplo, extraer datos de facturas o registrar información en sistemas ERP). Si bien son útiles para flujos de trabajo predecibles, estas herramientas no están diseñadas para interpretar lenguaje natural ni realizar consultas semánticas, lo cual limita su aplicabilidad en escenarios más dinámicos (Davenport & Ronanki, 2018).

### **4. Sistema inteligente con IA, OCR, LLM y bases vectoriales (solución propuesta)**

La alternativa seleccionada consiste en un sistema que integra reconocimiento óptico de caracteres (OCR), modelos de lenguaje de gran tamaño (LLM) y bases de datos vectoriales. Esta solución permite no solo digitalizar documentos, sino también analizarlos, interpretarlos y responder consultas sobre su contenido en lenguaje natural. Además, facilita búsquedas semánticas y clasificación automática, mejorando notablemente la eficiencia operativa.

La arquitectura modular basada en herramientas como Tesseract OCR, Streamlit, Qdrant y modelos como Llama 3.1 o Phi-3, ofrece flexibilidad, escalabilidad y bajo costo al usar recursos de código abierto. Esta alternativa se alinea con las proyecciones de Gartner (2022), que señalan que más del 50 % de las empresas adoptarán soluciones con IA para la gestión documental en los próximos años.

En consecuencia, la elección de esta solución responde a una combinación de criterios técnicos y estratégicos, ofreciendo una mayor capacidad de adaptación a diferentes sectores y garantizando precisión, trazabilidad y cumplimiento de normativas legales de protección de datos (Congreso de Colombia, 2012; European Union, 2016).

## **Análisis de restricciones**

Los proyectos enfocados en inteligencia artificial (IA) para la gestión documental, deben considerar múltiples restricciones que pueden limitar o condicionar su desarrollo. Estas restricciones incluyen factores tales como ambientales, económicos, legales, de salud y seguridad, socioculturales y técnicos, tanto externos como internos.

A continuación, se analizan cada una de ellas en el contexto del sistema propuesto.

### **Restricciones Ambientales**

Aunque los sistemas de gestión documental digitales no interactúan directamente con un entorno físico, se realiza el uso de servidores, bases de datos en la nube y modelos de IA, esto implica un consumo significativo de recursos energéticos. Según un informe de la Agencia Internacional de Energía (IEA, 2023), los centros de datos son responsables de aproximadamente el 1.5% del consumo eléctrico mundial. Cifra que sin lugar a duda ha incrementado estos últimos años ya que el uso e implementación de IA en todos los entornos se encuentra en auge.

En este sentido, es importante analizar el posible uso de proveedores que utilicen fuentes de energía renovable y gestionen la eficiencia energética conforme a estándares internacionales.

Además, se deben revisar los Planes de Ordenamiento Territorial (POT) y las licencias ambientales relacionadas con la ubicación física de servidores, si se gestionan localmente (Ministerio de Ambiente y Desarrollo Sostenible, 2021).

## **Restricciones Económicas**

En cuanto a las restricciones económicas del sistema depende de la disponibilidad de recursos financieros para su desarrollo, implementación y mantenimiento. El uso de tecnologías propietarias, el acceso a modelos de lenguaje grandes (LLM) y los costos asociados al almacenamiento y procesamiento de datos ya sea en un entorno local o en la nube puede elevar significativamente el presupuesto (OECD, 2021).

También se debe tener muy presente los factores macroeconómicos tales como cambios en políticas tributarias, subsidios a la tecnología, o escenarios de expropiación estatal pueden afectar la inversión (DANE, 2022).

## **Restricciones Legales**

En Colombia, la gestión de documentos digitales está regulada por normativas específicas como la Ley 594 de 2000, la Ley 1581 de 2012 sobre protección de datos personales, y su decreto reglamentario 1377 de 2013. Cada una de estas normas establecen lineamientos para el tratamiento, conservación, y disposición final de documentos, así como para la protección de datos sensibles. La solución de IA debe contar con funciones de trazabilidad, autenticación, y control de acceso para cumplir con estas disposiciones (Congreso de Colombia, 2000).

## **Restricciones de Salud y Seguridad**

Aunque el sistema no conlleva riesgos físicos, sí debe garantizar la ciberseguridad de la información. Tema que toma total relevancia ya que el activo más importante para una compañía son los datos, un mal manejo de accesos o una filtración de datos puede comprometer la integridad organizacional o poner en riesgo a los usuarios. Según el

Instituto Nacional de Ciberseguridad (INCIBE, 2022), así como también se deben cumplir estándares como el ISO/IEC 27001, que establece los requisitos para sistemas de gestión de la seguridad de la información.

## **Alternativa de solución**

El análisis financiero del proyecto contempla tres categorías principales: costos de operación, inversión inicial y capital de trabajo. Este enfoque permite estimar de forma detallada los recursos requeridos para el diseño, desarrollo, implementación y puesta en marcha del sistema inteligente de gestión documental. A continuación, se describen los rubros más relevantes.

### **1. Costos de operación**

Esta categoría incluye los gastos asociados al funcionamiento del sistema durante su fase activa. El componente principal es la mano de obra técnica, estimada en \$4.000.000 COP mensuales, correspondiente a un desarrollador full-stack con conocimientos en inteligencia artificial. Adicionalmente, se consideran los servicios en la nube (AWS o GCP), con un costo aproximado de \$2.400.000 COP por un periodo de seis meses.

Es importante destacar que el uso de modelos de lenguaje de código abierto como Llama 3.1, Phi-3 y Gemma no representa un costo adicional, al igual que las herramientas de desarrollo complementarias (VS Code, GitHub Free), lo cual reduce significativamente la carga financiera del proyecto. Se contempla también un valor de \$1.400.000 COP para publicidad, soporte básico y legal inicial. El subtotal de operación asciende a \$7.800.000 COP.

### **2. Inversión inicial**

Esta categoría contempla los elementos indispensables para arrancar el proyecto. Se incluye la adquisición de un equipo de desarrollo con especificaciones de alto

rendimiento (ThinkStation con RTX A2000, SSD y DDR5), cuyo valor total es de \$10.424.400 COP (incluyendo impuestos).

También se contemplan los costos de instalación y configuración de herramientas tecnológicas, estimados en \$1.600.000 COP, y el diseño técnico mediante wireframes y pruebas preliminares, valorado en \$1.400.000 COP. La curaduría y anotación de datasets para pruebas asciende a \$1.200.000 COP, mientras que los costos por certificados y normativas fueron eliminados por tratarse de un uso académico. El subtotal de inversión inicial es de \$14.624.400 COP.

### **3. Capital de trabajo**

En esta categoría se incluye la nómina de soporte técnico post-pruebas, correspondiente a un mes de trabajo de un desarrollador y un asistente para mantenimiento, con un costo de \$4.800.000 COP. Además, se asigna un fondo de contingencia de \$1.000.000 COP para ajustes legales o técnicos menores. El subtotal de capital de trabajo es de \$5.800.000 COP.

### **Total estimado del proyecto**

Sumando las tres categorías anteriores, el costo total estimado del proyecto asciende a \$28.224.400 COP. Esta estimación evidencia un uso eficiente de los recursos, priorizando el uso de herramientas open source y tecnología ya adquirida, lo cual permite optimizar la inversión sin comprometer la calidad ni los objetivos técnicos del sistema.

Categoría	Elemento	Descripción	Costo estimado (COP)
<b>1. Costos de operación</b>	Mano de obra técnica	Desarrollador full-stack + Especialista IA (ajustado)	\$4.000.000
	Servicios en la nube	AWS o GCP para almacenamiento + ejecución (6 meses)	\$2.400.000
	Modelos LLM (open source)	Uso de modelos como Llama 3.1, Phi-3, Gemma — sin costo de licencia	\$0
	Licencias software complementario (open)	Uso de herramientas gratuitas (VSCode, GitHub Free, etc.)	\$0
	Publicidad y administración (Overhead)	Página web, redes, soporte básico, legal inicial	\$1.400.000
<b>Subtotal operación</b>			<b>\$7.800.000</b>
<b>2. Inversión inicial</b>	Equipo de desarrollo (ThinkStation)	Ya adquirido – RTX A2000, SSD, DDR5 (USD 2.190 + IVA 19%)	\$10.424.400
	Instalación y configuración de herramientas	Configuración inicial OCR/LLM/vector DB	\$1.600.000
	Diseño y arquitectura técnica	Wireframes, UML, pruebas iniciales	\$1.400.000
	Dataset y documentación para pruebas	Documentos simulados, curaduría y anotación	\$1.200.000
	Certificados y normativas	Eliminado (uso académico)	\$0
<b>Subtotal inversión</b>			<b>\$14.624.400</b>
<b>3. Capital de trabajo</b>	Nómina soporte (1 mes post-pruebas)	Un desarrollador + asistente para mantenimiento	\$4.800.000
	Fondo de contingencia / ajustes	Ajustes legales/técnicos menores	\$1.000.000
<b>Subtotal capital trabajo</b>			<b>\$5.800.000</b>
<b>Total estimado del proyecto</b>		Sumatoria de las tres categorías	<b>\$28.224.400</b>

*Tabla 1. Análisis de costos, creación propia*

## **Plan de Implementación del Sistema**

La implementación del sistema propuesto se estructura en cinco fases progresivas que garantizan la planificación técnica, la ejecución eficiente de recursos y la validación funcional del prototipo. Este plan tiene como objetivo asegurar la entrega de un sistema funcional y adaptable a los entornos reales, con mínimo margen de error y máxima utilidad para los usuarios finales.

### **Fase 1: Preparación y alistamiento técnico**

#### **Actividades clave:**

Revisión de requerimientos funcionales y no funcionales.

Instalación de entorno de desarrollo (Python, VSCode, GitHub).

Configuración de infraestructura en la nube (AWS o GCP).

### **Fase 2: Desarrollo del prototipo**

#### **Actividades clave:**

Desarrollo del frontend en Streamlit.

Integración del módulo OCR (Tesseract, Google Vision API).

Configuración de base de datos vectorial (Qdrant o FAISS).

Conexión con modelos LLM (Llama 3.1, Phi-3).

Pruebas internas de funcionamiento y corrección de errores iniciales.

### **Fase 3: Pruebas piloto y ajuste**

#### **Actividades clave:**

Ejecución de pruebas con documentos reales (PDF, imágenes, Word).

Evaluación de precisión del OCR y tiempo de respuesta.

Ajustes técnicos en los modelos y estructura de datos.

Validación de usabilidad con usuarios de prueba.

#### **Fase 4: Documentación y capacitación**

##### **Actividades clave:**

Elaboración de manual técnico del sistema.

Redacción de la guía de usuario para uso no técnico.

Capacitación básica a stakeholders sobre el uso del sistema.

#### **Fase 5: Cierre, socialización y entrega final**

##### **Actividades clave:**

Presentación de resultados del proyecto a los stakeholders.

Revisión final del sistema y documentación.

Entrega oficial del prototipo funcional.

Recomendaciones para mejoras futuras o escalabilidad.

## **Conclusiones**

A través del uso de modelos de lenguaje de código abierto (como Llama 3.1 y Phi-3), herramientas de procesamiento OCR gratuitas (como Tesseract) y bases de datos vectoriales como Qdrant, el proyecto demuestra que es posible construir un sistema funcional de gestión documental sin incurrir en altos costos de licencias, lo cual lo hace viable para pequeñas y medianas organizaciones

La solución aborda problemas reales como la pérdida de productividad y la dificultad para acceder a información en formatos no estructurados. Su aplicación potencial en sectores como salud, legal y financiero garantiza un alto impacto, al mejorar la precisión en la consulta, reducir el tiempo de búsqueda de información y fortalecer el cumplimiento normativo en la gestión documental.

Con un costo total estimado de \$28.224.400 COP, el prototipo logra un equilibrio entre funcionalidad, eficiencia operativa y viabilidad económica. Esto permite considerar futuras fases de validación en entornos reales o incluso su evolución hacia un producto SaaS, con posibilidades de monetización y retorno de inversión en un horizonte a corto o mediano plazo

## Referencias

- Agencia Internacional de Energía [IEA]. (2023). *Data centers and data transmission networks*. <https://www.iea.org/reports/data-centres-and-data-transmission-networks>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
- Bharadiya, J., Thomas, A., & Ahmed, S. (2023). *Advances in OCR for document digitization*. Springer.
- Bharadiya, J. P., Thomas, R. K., & Ahmed, F. (2023). Rise of artificial intelligence in business and industry. *Journal of Engineering Research and Reports*, 25(3), 85–103.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 1877–1901.
- Cengel, Y. A., & Boles, M. A. (2019). *Termodinámica* (8.<sup>a</sup> ed.). McGraw-Hill.
- Congreso de Colombia. (2000). *Ley 594 de 2000 - Ley General de Archivos*. <https://www.funcionpublica.gov.co>
- Congreso de Colombia. (2012). *Ley 1581 de 2012 - Protección de datos personales*. <https://www.funcionpublica.gov.co>

DANE. (2022). *Contexto macroeconómico nacional y tendencias económicas 2022*.  
<https://www.dane.gov.co>

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world.  
*Harvard Business Review*, 96(1), 108–116.

Enholt, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, 24(5), 1709–1734.

European Union. (2016). *General Data Protection Regulation (GDPR)*. <https://gdpr.eu>

Gartner. (2022). *Top 10 strategic technology trends for 2025*. <https://www.gartner.com>

González, L., Pérez, D., & Ramírez, J. (2020). *Gestión de proyectos de ingeniería: una perspectiva integral*. Ediciones Universidad Nacional.

Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020). REALM: Retrieval-Augmented Language Model Pre-Training. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*.

Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.

Hernández, R., Fernández, C., & Baptista, P. (2020). *Metodología de la investigación* (6.<sup>a</sup> ed.). McGraw-Hill.

Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172.

IBM. (2018). *The cost of inefficient document management*. <https://www.ibm.com>

INCIBE. (2022). *Normas internacionales de seguridad de la información*.  
<https://www.incibe.es>

Long, S., He, X., & Yao, C. (2021). Scene text detection and recognition: The deep learning era. *International Journal of Computer Vision*, 129(1), 161–184.

McKinsey & Company. (2016). *The social economy: Unlocking value and productivity through social technologies*. McKinsey Global Institute.

Ministerio de Ambiente y Desarrollo Sostenible. (2021). *Lineamientos para la evaluación ambiental estratégica en planes de ordenamiento territorial*.  
<https://www.minambiente.gov.co>

Ministerio TIC. (2023). *Brecha digital y capacidades institucionales en Colombia*.  
<https://www.mintic.gov.co>

OECD. (2021). *AI in the digital economy: Implications for productivity and growth*.  
<https://www.oecd.org>

OpenAI. (2023). *GPT-4 technical report*. *arXiv preprint arXiv:2303.08774*.

Patel, M., Sharma, R., & Gupta, P. (2021). Deep learning-based optical character recognition: Trends and applications. *Journal of Artificial Intelligence Research*, 70, 123–145.

Pérez, J. A., & Vargas, M. L. (2021). *Inteligencia artificial aplicada a la gestión documental: Casos y aprendizajes en América Latina*. Universidad del Valle.

Pressman, R. S., & Maxim, B. R. (2020). *Software engineering: A practitioner's approach* (9th ed.). McGraw-Hill.

Pugh, S. (1991). *Total design: Integrated methods for successful product engineering*. Addison-Wesley.

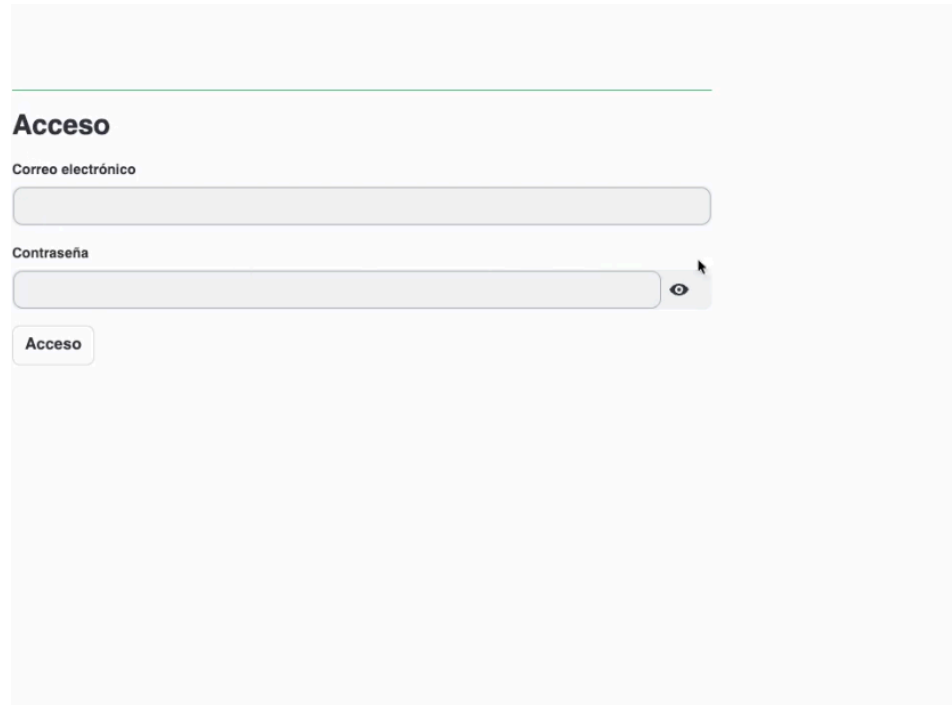
Pure Storage. (2025). *¿Qué es la administración de datos no estructurados?*  
<https://www.purestorage.com/la/knowledge/what-is-unstructured-data-management.html>

Smith, R. (2007). An overview of the Tesseract OCR engine. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR)*, 629–633.

Su, B., Lu, S., & Tan, C. L. (2014). Character recognition in natural scenes: A survey. *Pattern Recognition*, 48(3), 1033–1045.

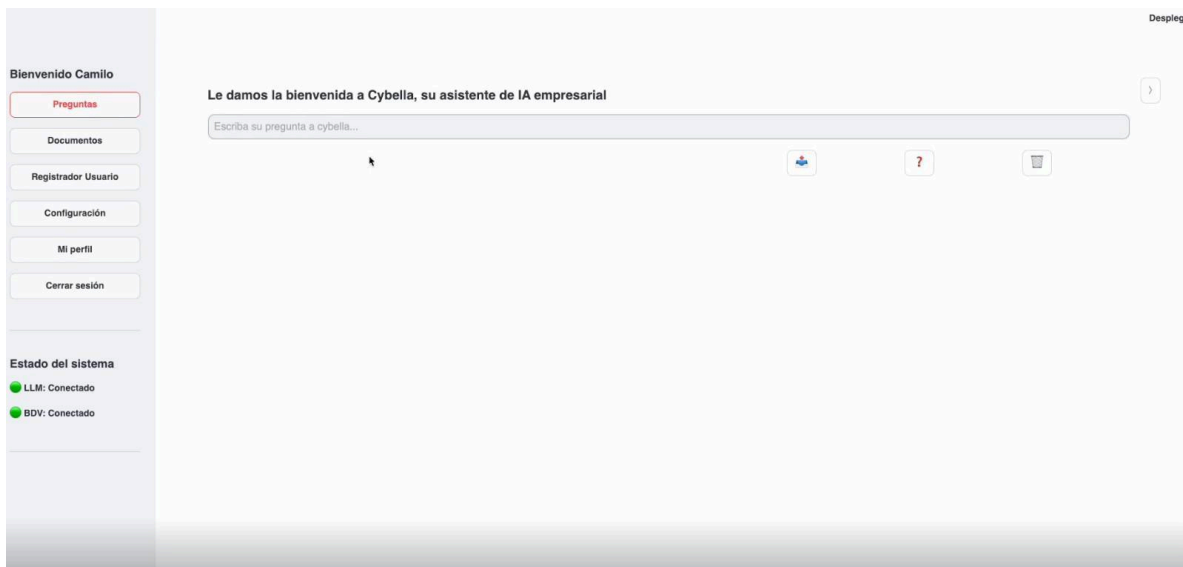
Treveil, A., Woolf, J., James, S., Turner, J., & Kahn, S. (2020). *The AI-powered enterprise: Harness the power of ontologies to make your business smarter, faster and more profitable*. O'Reilly Media.

## Anexos



A login form prototype with a light gray background. At the top, the word "Acceso" is displayed in bold black text. Below it, there are two input fields: "Correo electrónico" and "Contraseña". The "Contraseña" field has a small eye icon to its right. At the bottom of the form, there is a button labeled "Acceso".

*Imagen 2. Login prototipo, creación propia*



A dashboard prototype with a light gray background. On the left side, there is a sidebar with the following elements: "Bienvenido Camilo", a "Preguntas" button (highlighted in red), "Documentos", "Registrador Usuario", "Configuración", "Mi perfil", and "Cerrar sesión". Below these is a section titled "Estado del sistema" with two green status indicators: "LLM: Conectado" and "BDV: Conectado". The main content area features a heading "Le damos la bienvenida a Cybella, su asistente de IA empresarial" and a large input field with the placeholder text "Escriba su pregunta a cybella...". To the right of the input field are three icons: a blue plus sign, a red question mark, and a gray trash can. In the top right corner, there is a "Despleg" button.

*Imagen 3. Prototipo, creación propia*

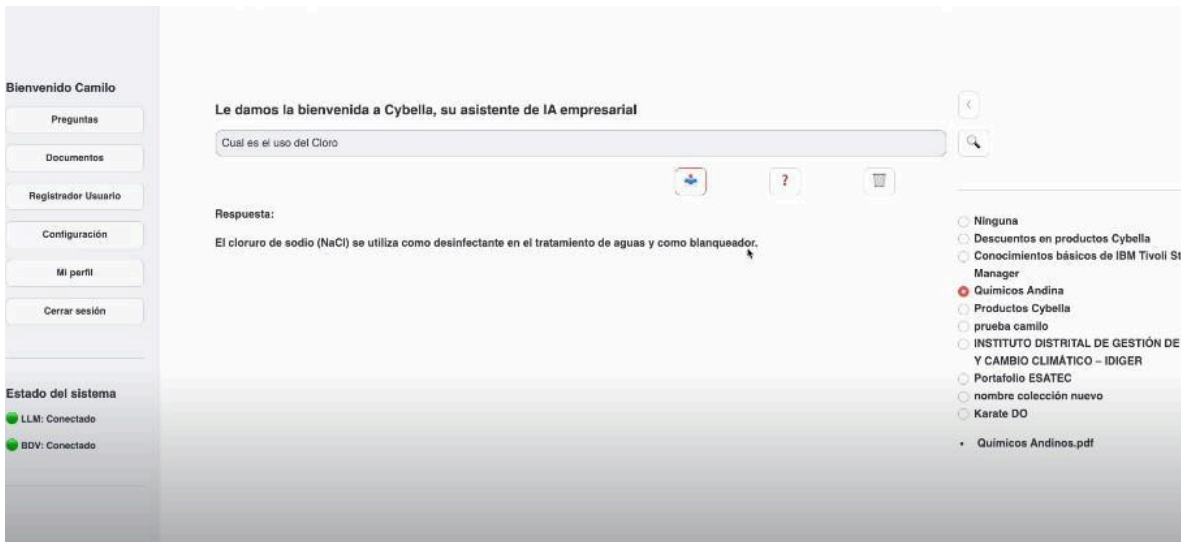


Imagen 4. Prototipo en ejecución, creación propia

## Modelo de LLM Actual

Modelo de mistral:

Se esta usando el modelo de LLM de Ollama Mistral:7b, en el cual se evidencio en base a las preguntas y respuestas:

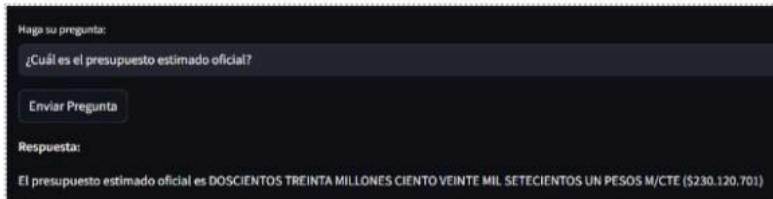
**Pregunta: ¿Cuáles pueden ser las causas de suspensión del contrato?**



\* Tiempo de respuesta: 2 minutos

**Respuesta que concuerda**

**Pregunta: ¿Cuál es el presupuesto estimado oficial?**



\* Tiempo de respuesta: 1 minuto y 34 segundos

**Respuesta que concuerda**

Imagen 5. Pruebas de tiempos con llm, Creación propia

