



Soluciones de IA Generativa

La Pampa - Relevamiento, análisis, recomendaciones y sugerencias - Versión 1.0

Contenido

1	Introducción	3
2	Relevamiento Detallado	3
2.1	Mapeo completo de procesos administrativos y técnicos	3
	Inventario de fuentes de datos existentes, formato y calidad	4
	Evaluación de accesibilidad de datos (sincrónica/asincrónica)	6
	Recolección de datos de muestra y evaluación de compatibilidad	6
3	Reuniones de Cierre del Relevamiento	7
	Presentación de hallazgos preliminares	7
	Identificación de casos de uso con mayor factibilidad y potencial impacto	7
4	Análisis de Soluciones	8
4.1	Selección de tecnologías de IA generativa adecuadas	8
	Evaluación de herramientas y plataformas disponibles	9
	Diseño técnico preliminar para soluciones propuestas	10
	Análisis de viabilidad técnica y operativa	11
5	Prueba de Concepto	12
5.1	Desarrollo de la prueba de concepto	12
	Definición del conjunto de datos a utilizar	12
	Ensayos preliminares	12
	Diseño e implementación del prototipo funcional	12
	Despliegue del entorno de prueba y validación funcional	13
	Ajustes y optimización del prototipo	13
	Recomendaciones para implementación futura	14

1 Introducción

El presente documento recopila los procesos relevados junto con la Dirección General del Registro de la Propiedad Inmueble (RPI) y la Dirección General de Superintendencia de Personas Jurídicas y Registro Público de Comercio (DGSPJyRPC) con potencial de ser agilizados mediante el uso de Inteligencia Artificial Generativa.

En el marco de este proyecto se plantean posibles enfoques para optimizar dichos procesos, se documentan los principales desafíos de cada uno, y se ensaya una prueba de concepto para analizar el impacto posible de la solución.

2 Relevamiento Detallado

2.1 Mapeo completo de procesos administrativos y técnicos

Mediante reuniones regulares realizadas con ambos registros, se llevó a cabo un relevamiento para comprender mejor el proceso y las potenciales oportunidades de optimización.

Registro de la Propiedad (DGRPI)

- El proceso inicia en una **mesa de entrada única** donde cada trámite recibe un **número de entrada/año** que define su prioridad ("primero en el tiempo, primero en el derecho").
- El ingreso lo realiza un profesional externo, mediante **formulario físico** y testimonio firmado.
- La información de cada trámite se ingresa al sistema **MAPUN**, que tiene aprox. el 40% de los inmuebles cargados.
- El proceso incluye: verificación del formulario con la escritura → carga en oficina de Folio → generación de minuta digital.
- Se verifica que no haya doble asiento (solo en papel o digital).
- Las registraciones contienen: datos del inmueble, titularidad, gravámenes, antecedentes y estado del trámite.
- Alta carga de trabajo: 60% compra/venta, 80% notariales, 90% judiciales (sucesiones/divorcios).

Flujo

1. Ingreso de trámite → numeración → vinculación con profesional.
2. Carga al sistema MAPUN → clasificación por tipo → derivación a oficinas.
3. Verificación (operador) de consistencia entre escritura, formulario y antecedentes.
4. Carga en Folio → validación → sellado → minuta digital.

Personas Jurídicas (DGSPJyRPC)

- Utilizan un sistema centralizado que consume datos desde **RENAPER**.
- Escanean documentos (estatutos, actas, etc.), los almacenan en un **servidor físico compartido**.
- Documentos manuscritos digitalizados con baja calidad, especialmente anteriores a 1990.
- Reciben archivos adicionales como Excel con nómina de autoridades.
- Objetivo: **reducir papel y actualizar procesos**.
- En algunos casos, no envían toda la información (p.ej. mandan folio pero no antecedentes completos)

Flujo

1. Ingreso de documentación (formato manuscrito o digital).
2. Escaneo → almacenamiento en servidor físico.
3. Carga parcial de datos en el sistema desde RENAPER o Excel (según el caso).
4. Comparación manual entre Excel, escaneo y expediente.

Inventario de fuentes de datos existentes, formato y calidad

Ambos registros compartieron muestras e información sobre los distintos tipos de documentos involucrados en cada uno de los trámites y tareas que realizan con frecuencia.

Registro de la Propiedad (DGRPI)

- **Formularios físicos de ingreso**
 - Formato: papel (escaneado a PDF, como imagen embebida)

- Calidad: buena, estructurada, escaneados al ingresar

- **Minutas de escrituras**

- Formato: papel/PDF (como imagen embebida)
- Calidad: variable según antigüedad, pero en general legible

- **Sistema MAPUN**

- Formato: digital, estructurado
- Calidad: alto, pero cobertura parcial (~40% de inmuebles)

- **Antecedentes históricos**

- Formato: papel (tomos antiguos), parte digitalizada
- Calidad: variable; tomos antes de 1950, digitalización incompleta

- **Datos del RENAPER**

- Formato: digital (API)
- Calidad: alta confiabilidad

Personas Jurídicas (DGSPJyRPC)

- **Documentos escaneados (estatutos, actas, etc.)**

- Formato: PDF (escanearon manuscritos, imágenes embebidas)
- Calidad: baja en documentos anteriores a 1990

- **Excel con nómina de autoridades**

- Formato: digital (Excel)
- Calidad: alta, estructurada

- **Datos del RENAPER**

- Formato: digital (API), acceso manual actualmente
- Calidad: alta

- **Expedientes históricos (folios, tomos)**

- Formato: papel, algunos digitalizados
- Calidad: variable, incompleto

- **Instructivos / Marco normativo**

- Formatos: PDF (gran mayoría en texto, algunos como imagen)
- Calidad: Alta, salvo excepciones puntuales

- **Modelos de notas**

- Formato: Documento de Microsoft Word
- Calidad: Alta

Evaluación de accesibilidad de datos (sincrónica/asincrónica)

Registro de la Propiedad (DGRPI)

- **MAPUN**: acceso **asíncrono**, requiere procesamiento interno
- **Formularios y minutas físicas**: acceso **manual**, luego escaneo
- **RENAPER (verificación de identidad)**: acceso **sincrónico** mediante API
- **Antecedentes en papel**: acceso **manual y demorado**

Personas Jurídicas (DGSPJyRPC)

- **Servidor con escaneos**: acceso **asíncrono**, en servidor físico compartido
- **Excel (nóminas)**: acceso **manual**
- **RENAPER**: acceso **sincrónico** (manual hoy, puede automatizarse)
- **Expedientes físicos**: acceso **manual**, algunos digitalizados parcialmente

Recolección de datos de muestra y evaluación de compatibilidad

Registro de la Propiedad (DGRPI)

- Se sugiere recolectar una muestra de formularios y escrituras que ya hayan sido ingresados al sistema, y que hayan pasado por todas las etapas.
- Evaluar la compatibilidad entre los 13 rubros clave (de la escritura, formulario y antecedentes) para ver si la IA puede automatizar la validación cruzada.????

Personas Jurídicas (DGSPJyRPC)

- Recolectar un set de escaneos manuscritos, sus respectivas versiones en Excel y el expediente final.
- Digitalizar el texto y analizar con herramientas de NLP si es posible automatizar el cruce de datos (nómina, autoridad, estatutos).
- Evaluar diferencias entre asociaciones, sociedades y diferentes entes.

3 Reuniones de Cierre del Relevamiento

Presentación de hallazgos preliminares

Se trabajó en conjunto con la Dirección General del Registro de la Propiedad Inmueble y la Dirección General de Superintendencia de Personas Jurídicas y Registro Público de Comercio. Ambas entidades poseen gran cantidad de información respaldada en soporte papel digitalizada, si bien ha sido mencionado en diversas reuniones que se encuentran en un proceso de transformación digital para algunos de sus trámites.

El procesamiento digital de los documentos actualmente sólo comprende el escaneo de las hojas, lo que dificulta su búsqueda y/o interpretación por parte de cualquier sistema informático.

Ambas direcciones, a su vez, poseen información publicada en sus respectivas páginas web:

- DGRPI
 - https://mapun.lapampa.gob.ar/mapun/servlet/com.rpi.consultanormativa_wl
- DGSPJyRPC
 - <https://dgpersonasjuridicas.lapampa.gob.ar/sociedades-1.html>
 - <https://dgpersonasjuridicas.lapampa.gob.ar/entidades-civiles-1.html>
 - <https://dgpersonasjuridicas.lapampa.gob.ar/marco-normativos.html>

Esta información, si bien se encuentra disponible, no posee ningún tipo de buscador para poder acceder fácilmente a la información solicitada.

En reuniones, tanto conjuntas como con cada dirección en particular, se validó la necesidad y se debatieron potenciales enfoques para agilizar y simplificar diversos trámites.

Identificación de casos de uso con mayor factibilidad y potencial impacto

Se ha determinado que los siguientes escenarios son pasibles de aplicación de soluciones de IA Generativa:

Registro de la Propiedad (DGRPI)

- Extracción de datos relevantes de escrituras y validación contra formulario

Como parte de la solicitud de registración, los profesionales suelen enviar un formulario que describe las partes involucradas en el trámite, montos, inmueble afectado, y certificados, entre otros. Además, presentan la escritura del inmueble. A los efectos de llevar a cabo la solicitud, en el Registro de la Propiedad Inmueble se contrastan estos datos.

Personas Jurídicas (DGSPJyRPC)

- Interfaz conversacional automatizada (chatbot) para resolución de dudas de trámites

La página web de la DGSPJyRPC cuenta con documentos detallando los distintos trámites y documentación que debe presentarse para cada uno de ellos. Los profesionales tienen acceso a ellos manualmente, debiendo dedicar tiempo hasta dar con la información requerida.

- Indexación inteligente de documentos escaneados

Las actas de constitución y modificación de sociedades poseen respaldo en soporte papel. Las mismas, como parte de un proceso de transformación, han sido digitalizadas como imagen, sin ningún tipo de interpretación de sus datos. Al necesitar acceder a un libro específico, se debe realizar una búsqueda manual, ya que la información relevante no se encuentra indexada.

- Conformación de historial de sociedad basado en actos jurídicos previos

Como parte del documento de precalificación, se confecciona un documento detallando el acta de constitución de la sociedad, y sus posteriores actas relevantes.

4 Análisis de Soluciones

4.1 Selección de tecnologías de IA generativa adecuadas

La tecnología de Inteligencia Artificial generativa que mejor se ajusta para los casos de uso propuestos, tanto del RPI como de la DGSPJyRPC, es de tipo LLM (Large Language Model - Gran modelo de lenguaje), que permite vía *prompts* (instrucciones de texto) realizar diversas tareas a demanda con los datos. Asimismo, para algunas tareas accesorias donde la imagen sea un factor relevante en la resolución, se puede

utilizar un VLM (Vision Language Model - Modelo de lenguaje de visión) que añade la posibilidad de procesar imágenes, obteniendo como resultado una salida de texto con lo solicitado. No obstante, este tipo de modelos no suelen ser adecuados para tareas de obtención de texto a partir de imágenes, cuestión de vital importancia para aquellos documentos escaneados.

Evaluación de herramientas y plataformas disponibles

- Extracción de datos relevantes de escrituras y validación contra formulario (RPI) e Indexación inteligente de documentos escaneados (DGSPJyRPC)

Para obtener la información estructurada de los documentos, como por ejemplorazón social, personas involucradas, fecha de inscripción, etc. se pueden utilizar LLMs a fin de solicitar la información a extraer. Sin embargo, esta tarea involucra también algunas técnicas complementarias como OCR (Optical Character Recognition - Reconocimiento óptico de caracteres) y la recolección y organización de los documentos en una base de datos.

Para tal fin, si bien hay en el mercado un amplio abanico de herramientas de orquestación de procesos como Make y n8n, consideramos más apropiado para ensayar una prueba de concepto, tomar como base una herramienta que ya hemos desarrollado internamente, que no es más que una plataforma orientada a la clasificación de documentos y a la extracción estructurada de datos de los mismos.

En dicha plataforma hemos vertido nuestra vasta experiencia en el procesamiento de documentos, con la posibilidad de desarrollar pequeños módulos complementarios para solucionar desafíos puntuales. Esto permite de forma ágil componer soluciones incluso ante escenarios donde no sea posible establecer desde el inicio los distintos tipos de documentos posibles y la información de interés a extraer de ellos, ya que al ser modular y configurable es posible adaptarla a los distintos casos de borde que puedan surgir, situación esperable en cualquier proyecto que involucre datos.

Otra alternativa es desarrollar una solución ad-hoc para este fin, pero considerando que la tarea clave a realizar posee un gran grado de solapamiento con esta plataforma, implicaría terminar desarrollando desde cero una plataforma similar, con el consiguiente tiempo que esto conlleva.

- Interfaz conversacional automatizada (chatbot) para resolución de dudas de trámites (DGSPJyRPC)

Existen varias herramientas hoy en día enfocadas en el diseño de flujos conversacionales. Una de ellas es LangGraph, que permite diseñar grafos ante distintos escenarios y consultas por parte del usuario. En nuestra experiencia, estos *frameworks* (estructuras prearmadas en torno a una librería) agilizan en buena

medida el desarrollo de este tipo de aplicaciones, pero suelen ser inconvenientes si se requiere alguna funcionalidad no contemplada por el desarrollador de la librería.

Por otra parte, se debe realizar la ingesta de los datos desde la página web de la DGSPJyRPC, lo que incluye documentos en PDF tipo imagen. Aquí se torna necesario el uso de un proceso de OCR, además de un motor de búsqueda capaz de encontrar las secciones más relevantes de texto con el objetivo de conformar una respuesta al usuario. Para esta última tarea, una herramienta capaz de ello es Elasticsearch, que permite realizar búsqueda vectorial. Con esta técnica los vectores que se asignan a cada sección de documento poseen una codificación semántica, es decir, que condensan el significado del pasaje de texto en una serie de números de longitud fija. Al mismo tiempo, se realiza una búsqueda tradicional por palabras clave, con el fin de capturar de forma amplia resultados que enriquezcan la respuesta al usuario.

Una vez indexada esta información, será un agente conversacional el encargado de confeccionar el mensaje mostrado. Los agentes son automatizaciones que apoyan sus decisiones en consultas a un LLM, donde pueden optar por utilizar herramientas que les son provistas (por ejemplo, para acceder o modificar una base de datos) o bien elaborar una respuesta al usuario.

- Conformación de historial de sociedad basado en actos jurídicos previos (DGSPJyRPC)

En base a las extracciones realizadas se puede orquestar un agente que complete la información más relevante con respecto a una sociedad en base a la información indexada en la base de datos. Además de indexar las entidades más relevantes de cada documento, puede confeccionarse un resumen del contenido. Este proceso puede llevarse a cabo con una herramienta de orquestación como las anteriormente mencionadas, como también encuadrarse dentro de un desarrollo más personalizado. Es necesario tener presente que al optar por herramientas prearmadas de orquestación siempre se tienen limitaciones técnicas que en ocasiones pueden llegar a dificultar la obtención de los resultados esperados.

Diseño técnico preliminar para soluciones propuestas

- Extracción de datos relevantes de escrituras y validación contra formulario (RPI) e Indexación inteligente de documentos escaneados (DGSPJyRPC)

Para la extracción de datos de escrituras y su validación contra el formulario, un posible enfoque es realizar una extracción de datos estructurada tal que la salida para ambos casos (escrituras y formulario) posean campos idénticos, o lo más similares posibles en su defecto. Esta tarea se llevaría a cabo con una plataforma similar a nuestra herramienta interna, dada la facilidad para configurar distintos esquemas de

extracción. Luego, ambas estructuras serán comparadas mediante el empleo de un LLM, el cual realizará la comparación de forma completa o por secciones. Queda abierta la integración de esta solución al sistema MAPUN en caso de ser necesario, con lo cual se expondría una API REST debidamente documentada para tal fin.

En el caso de la indexación inteligente, se debe utilizar un VLM para determinar de manera acertada dónde comienzan y terminan los documentos. Una vez realizada esta separación, se llevará a cabo la extracción de información estructurada utilizando una herramienta similar a nuestro desarrollo, previa detección del tipo de documento que se trate con la misma herramienta. Una vez obtenida la información, se procederá a almacenar esta información en el motor de búsqueda Elasticsearch (o una base de datos ya preexistente).

Tal como en el caso anterior, queda abierta la posibilidad de integración mediante una API REST documentada, o bien incluir el desarrollo de una interfaz web para su uso y administración.

- Interfaz conversacional automatizada (chatbot) para resolución de dudas de trámites (DGSPJyRPC)

El chatbot se llevaría a cabo sin librerías que posean herramientas prearmadas para tener máxima flexibilidad a la hora de implementar los flujos conversacionales y las fuentes de datos a consultar.

La ingesta de los datos se realizará a partir de la página web del DGSPJyRPC, donde se encuentran múltiples archivos en PDF, algunos de los cuales presentan imágenes escaneadas. Tal como en los procesos anteriores que involucran archivos de similares características, se vuelve necesaria la utilización de un proceso de OCR.

Análisis de viabilidad técnica y operativa

En cuanto a la viabilidad técnica del proyecto, se posee conocimiento de todas las tecnologías que se estiman necesarias. Tanto el uso de tecnologías de IA generativa, como el desarrollo de arquitecturas escalables son claves para el éxito de la implementación de estas soluciones, especialmente teniendo en cuenta el volumen potencial de datos a procesar.

Asimismo, se asume suficiente para la integración con los sistemas actuales del RPI y la DGSPJyRPC la disponibilización de los servicios mediante una API REST.

Operativamente las soluciones propuestas se fundamentan en necesidades concretas y son pasibles de integrarse en el flujo de trabajo de ambos organismos. En nuestro enfoque ensayado, el evitar utilizar *frameworks* permite que la solución pueda

adaptarse en un futuro a más casos de uso, sin limitaciones técnicas que impliquen cambios de gran escala.

5 Prueba de Concepto

5.1 Desarrollo de la prueba de concepto

Definición del conjunto de datos a utilizar

Para el desarrollo de la aplicación de indexación inteligente se utilizaron actas escaneadas de la DGSPJyRPC correspondientes al año 2000. Las mismas poseen distintos tipos de documentos, entre los cuales se encuentran actas de constitución de sociedades, certificaciones de firmas, inventarios de iniciación de actividades, actas de inscripciones de directorios, balances contables y contratos de cesiones de cuotas.

En el caso del chatbot, se tomaron como referencia algunos documentos PDF con los requisitos para los trámites publicados en la página web de la DGSPJyRPC (<https://dgpersonasjuridicas.lapampa.gob.ar/>).

Ensayos preliminares

El potencial de implementación de la prueba de concepto se validó realizando una extracción sobre secciones del documento utilizando nuestra herramienta interna de extracción. Al cargar los documentos en la plataforma, se extrajeron entidades relevantes para el problema de acuerdo a lo esperado, independientemente del tipo de escritura (manuscrito/impreso). Dado que este es el paso más elemental y crítico para una implementación exitosa, considerando la capacidad de reconocer y estructurar información que actualmente es sólo accesible mediante la búsqueda manual del acta, se considera un primer paso exitoso para el avance hacia un prototipo funcional.

Adicionalmente, como dicha herramienta integra un proceso de OCR, se establece una base de desempeño para los documentos escaneados que también aplica para los documentos de idéntica naturaleza que describen los distintos trámites y requisitos en el ámbito de la DGSPJyRPC.

Diseño e implementación del prototipo funcional

Fueron implementadas las siguientes funcionalidades en una misma interfaz visual:

1. Chatbot sobre documentación de trámites de la DGSPJyRPC

Permite, mediante una interfaz conversacional, realizar diversas preguntas sobre diversos trámites y sus requisitos. En el marco de este prototipo, se realiza la carga manual de los documentos en PDF tipo texto, excluyendo aquellos que requieren de un proceso de OCR.

2. Búsqueda inteligente sobre libro de actas de la DGSPJyRPC

Permite cargar un PDF que contiene un libro de actas escaneado. El servicio interpreta y agrupa las páginas en los distintos documentos, los clasifica según su contenido y realiza una extracción acorde de sus datos utilizando la herramienta de extracción ya mencionada. Acto seguido se indexan los datos en Elasticsearch, quedando las distintas actas disponibles para su búsqueda a partir de un nombre, apellido, CUIT, o cualquier dato clave de acuerdo a la definición establecida para cada tipo de documento en específico. En el marco de este prototipo, no se admitirán más de 50 páginas, y el soporte estará dado para tres tipos de documentos: Acta de constitución de sociedad, Certificación de firmas, y tipo desconocido.

Despliegue del entorno de prueba y validación funcional

El entorno de prueba se encuentra desplegado en el siguiente enlace

Sitio: <https://la-pampa-gen-ai.asanteit.com.ar/>

Usuario: LaPampa-RPI/PJRPC

Contraseña: L4P4mp4-63n41

Se encontrará disponible para su uso hasta el 21/10/2025.

Ajustes y optimización del prototipo

Como parte de la integración de nuestra herramienta de extracción al prototipo, se ensayaron distintas configuraciones para poder obtener la información relevante.

Considerando que la solución se apoya en el uso de LLMs para realizar la extracción, la comprensión de la estructura de los distintos tipos de documentos permite elaborar descripciones de los campos a extraer más adecuadas.

También, como parte de la detección de inicio/final de documento para poder subdividir el libro de actas de forma automática, se utilizó el paquete de OCR paddleOCR en conjunto con un LLM. Si bien los resultados iniciales resultaron satisfactorios, se ensayó el uso de un VLM (Visual Language Model - Modelo de Lenguaje Visual) para la misma detección. En nuestras pruebas, este último enfoque representó una mejora en el desempeño de ese módulo.

Recomendaciones para implementación futura

Por tratarse de una prueba de concepto, naturalmente existen varias oportunidades de mejora. En nuestro análisis, consideramos que las siguientes tareas mejorarán el desempeño de cara a una implementación más orientada a producción:

- Optimización del OCR

El software utilizado para el OCR impacta directamente en la calidad de la información extraída, ya que todo dato incorrectamente relevado en esta etapa deja de estar disponible para su búsqueda.

- Optimización del algoritmo de separación de actas

El algoritmo de separación de actas se ejecuta de manera secuencial y sincrónica. Esto hace que cuanto más largo sea el documento a procesar, más tiempo de ejecución lleve esta etapa. De ejecutarse en paralelo, podría optimizarse el tiempo que demora este procesamiento.

- Procesamiento escalable

Dependiendo de la cantidad de usuarios concurrentes en la aplicación, podría ser necesario realizar búsquedas y/o ingesta de datos de forma simultánea.

- Mejoras en la interfaz de usuario

La interfaz de usuario presentada en la prueba de concepto resulta adecuada para desarrollo y evaluación. En un ámbito productivo