

GUÍA DE ESTUDIO

Antes de conectar un agente jurídico: revisar permisos y acciones

Una evaluación práctica para decidir qué puede leer, proponer y ejecutar un asistente conectado a documentos y correo.

Fuentes revisadas: 2026-10-21 · Edición de lectura y trabajo.

Qué podrás hacer

01. Elegir una tarea de organización con documentos públicos.
02. Crear la matriz leer, proponer y ejecutar.
03. Probar un archivo ficticio con una instrucción ajena al encargo y registrar la respuesta.

Empezar por una tarea pequeña y comprobable

Caso ficticio. Un consultorio quiere que un asistente organice documentos públicos y prepare una lista de asuntos para revisión. El proveedor ofrece además conectar correo, calendario y carpetas completas. La pregunta útil no es cuántas funciones incorpora, sino qué permisos necesita la tarea elegida y qué podría ocurrir si interpreta mal una instrucción. Esta nota propone un piloto acotado; no recomienda conectar expedientes reales para descubrir después sus límites.

OWASP describe la agencia excesiva como el riesgo asociado a funcionalidad, permisos o autonomía superiores a los necesarios. Su guía propone controles de autorización en los sistemas y revisión humana de acciones de impacto. Es una referencia técnica, no una norma jurídica colombiana ni una garantía de seguridad del producto que se esté evaluando.

Una matriz de permisos antes de una demostración comercial

Propuesta editorial: enumerar cada función del piloto en una fila y clasificarla como lectura, creación de borrador o ejecución. Para el ejemplo bastaría leer una carpeta preparada con documentos públicos y generar una lista local. La capacidad de borrar archivos, reenviar mensajes o compartir carpetas no aporta al resultado elegido. Debe permanecer deshabilitada si no existe una necesidad y una autorización separadas.

Preparar un conjunto de prueba que incluya un documento incompleto, una referencia desactualizada y una instrucción contradictoria escrita dentro de un archivo ficticio. Observar si el asistente distingue el contenido consultado de las instrucciones del responsable. El resultado aceptable no es obedecer cualquier frase que encuentre, sino señalar la contradicción y mantenerse dentro de la tarea y los

permisos definidos. Ningún ensayo debe contener secretos, credenciales ni información reservada.

Antes de ampliar el piloto, pedir una explicación concreta de los datos utilizados, conservación, acceso del proveedor, subprocesadores, registro de acciones y mecanismo de revocación. Revisar cómo se detiene una ejecución y cómo se recupera un borrador. Una demostración fluida no prueba esos controles. Si una función pretende enviar, publicar, radicar o modificar derechos, la revisión debe situarse antes de esa acción y mostrar al responsable el contenido y el destinatario exactos.

La decisión de adopción debe quedar vinculada a evidencia: qué tarea se probó, qué permisos se concedieron, qué errores aparecieron y qué límites continúan abiertos. Un piloto satisfactorio para organizar documentos públicos no demuestra idoneidad para valorar prueba o decidir asuntos jurídicos.

Llévalo a la práctica

Tu recorrido de trabajo

01. Elegir una tarea de organización con documentos públicos.
02. Crear la matriz leer, proponer y ejecutar.
03. Probar un archivo ficticio con una instrucción ajena al encargo y registrar la respuesta.

Resultado esperado

Una ficha de piloto con permisos mínimos, controles y criterios de suspensión.

Antes de darlo por terminado

01. La tarea no requiere datos reservados.
02. Los permisos corresponden a una necesidad concreta.
03. Las acciones externas tienen un responsable y revisión previa.

Método y alcance

Nota de análisis y ejercicio editorial propio. Los hechos y documentos se contrastan con las fuentes enlazadas; las recomendaciones no se atribuyen a una autoridad salvo indicación expresa.

Fuentes originales

01. OWASP LLM06:2025 · Excessive Agency

<https://genai.owasp.org/llmrisk/llm062025-excessive-agency/>

Guía técnica del proyecto OWASP; no ley

02. Resolución FGN 0-0077 de 2026

<https://www.fiscalia.gov.co/media-portalweb-prd/2026-Resolucion-0-0077.pdf>

Resolución institucional · Colombia

Continúa en la web

<https://alejoqqa.com/notas/evaluar-agente-juridico-permisos.html>

Consulta la versión web para actualizaciones posteriores a la fecha de revisión.