

Evaluación del Sistema RAG

Análisis de Pruebas: Trámites TEA en Andalucía

Ángel Manuel Pereira Rodríguez

13 de noviembre de 2025

Índice

1. Resumen Ejecutivo	2
2. Evaluación Detallada por Criterio	2
2.1. Pertinencia de los Documentos Recuperados	2
2.2. Claridad y Corrección de las Respuestas	2
2.3. Grado de Apoyo Práctico (Accionabilidad)	2
3. Documentación de Errores y Limitaciones	3
3.1. Errores Detectados	3
3.2. Limitaciones del Sistema	3
4. Propuestas de Mejora Futura	3
5. Anexo: Tabla Resumen de Evaluación	5

1 Resumen Ejecutivo

El análisis de los 15 casos de prueba (`resultados_prueba.json`) revela un rendimiento **sobresaliente** del sistema RAG. El sistema demuestra una alta capacidad para recuperar contextos pertinentes y generar respuestas claras, precisas y, lo que es más importante, accionables para el usuario.

Destaca especialmente en la gestión de trámites complejos (ej. valoración de discapacidad) y en la provisión de contactos clave (ej. Autismo Andalucía, teléfono de información de discapacidad). Un punto fuerte notable es su capacidad para **evitar la alucinación**: cuando no encontró información sobre 'beneficios fiscales', pivotó correctamente hacia 'ayudas y becas' y recomendó a un asesor fiscal, lo cual es un comportamiento ideal.

2 Evaluación Detallada por Criterio

A continuación, se desglosa la evaluación según los criterios solicitados.

2.1 Pertinencia de los Documentos Recuperados

La relevancia de los *chunks* recuperados es **consistentemente alta**. El sistema logró extraer los fragmentos de texto exactos que contenían la información solicitada, incluso cuando se trataba de detalles muy específicos.

- **Datos Específicos:** El sistema localizó con éxito importes monetarios (ej. '913,00 euros' y '400,00 euros' para becas), organismos oficiales ('Consejería de Inclusión Social, Juventud, Familias e Igualdad' para la discapacidad) y entidades clave ('Autismo Andalucía', 'Autismo Sevilla').
- **Contactos:** Recuperó con éxito números de teléfono (954 24 15 65, 900.55.55.64) y correos electrónicos (autismoandalucia@telefonica.net) cuando eran pertinentes.
- **Procesos:** En preguntas sobre procedimientos (ej. atención temprana), los *chunks* recuperados contenían la secuencia correcta de pasos (ej. 'pediatra de atención primaria' -> 'Unidades de Atención Temprana' -> 'Centros de Atención Temprana (CAIT)').

2.2 Claridad y Corrección de las Respuestas

La calidad de la generación de lenguaje es **excelente**. Las respuestas son:

- **Claras y Concisas:** Utilizan un lenguaje natural y fácil de entender, evitando la jerga administrativa siempre que es posible.
- **Correctas:** Toda la información generada en las respuestas está fielmente respaldada por la evidencia en los *chunks* recuperados. No se observó contradicción ni invención de datos.
- **Gestión de Negativos:** Como se mencionó, la respuesta a '¿Qué beneficios fiscales...?' es un ejemplo perfecto de robustez. El sistema admitió no tener información sobre 'beneficios fiscales' (impuestos) y reorientó al usuario hacia la información que sí poseía (becas y ayudas económicas), evitando una respuesta incorrecta.

2.3 Grado de Apoyo Práctico (Accionabilidad)

Este es el punto más fuerte del sistema. El grado de apoyo práctico es **muy alto**. Las respuestas no solo 'informan', sino que 'guían' al usuario.

- **Pasos Clave:** Preguntas como '¿Cómo solicitar adaptaciones curriculares?' o '¿Cómo acceder a la atención temprana?' se responden con una secuencia clara de acciones (ej. 'contactar con el departamento de orientación', 'acudir al pediatra').
- **Identificación de Actores:** El sistema identifica correctamente a quién debe dirigirse el usuario: el 'Equipo de Orientación Educativa especializado en TEA' para adaptaciones, la 'Delegación Territorial' para la discapacidad, o 'Autismo Andalucía' para orientación general.
- **Resolución de Dudas:** La respuesta sobre la 'renovación' del certificado de discapacidad explica correctamente que el proceso es 'de oficio' y proporciona el teléfono de contacto para incidencias (pérdida o robo), lo cual es la información práctica que el usuario realmente necesita.

3 Documentación de Errores y Limitaciones

3.1 Errores Detectados

Durante la revisión de los 15 casos de prueba, ****no se han detectado errores fácticos ni alucinaciones**** en las respuestas generadas.

La única 'imprecisión' observada fue en la respuesta sobre 'centros en Sevilla', donde el sistema menciona '24 aulas específicas' pero no puede listarlas. Sin embargo, esto no es un error del sistema, sino una limitación de los datos fuente. El sistema manejó esta situación correctamente al redirigir al usuario a 'Autismo Sevilla', la entidad local pertinente para obtener dicho listado.

3.2 Limitaciones del Sistema

A pesar de su alto rendimiento, el sistema presenta limitaciones inherentes al enfoque RAG:

- **Dependencia Documental:** El sistema solo puede responder con la información contenida en su base de datos vectorial (FAISS). Si un nuevo trámite o una ley no ha sido *scrapeada* e indexada, el sistema no podrá informar sobre ello.
- **Normativa Cambiante:** Los importes de las becas, los nombres de las consejerías y los procedimientos pueden cambiar anualmente. La base de datos actual (basada en *scrapping* en una fecha concreta) quedará obsoleta.
- **Ausencia de Asesoramiento Personalizado:** El sistema proporciona información general de alta calidad, pero ****no puede sustituir el asesoramiento profesional**** de un trabajador social, un orientador educativo o un asesor jurídico que evalúe un caso particular.

4 Propuestas de Mejora Futura

Basado en las limitaciones detectadas, se proponen las siguientes líneas de mejora:

1. **Pipeline de Actualización Automatizada:** Implementar un *script* (ej. GitHub Action) que ejecute el *pipeline* de *scraping* y vectorización de forma periódica (ej. trimestralmente) para mantener la base de datos actualizada.
2. **Metadatos de Fecha en Chunks:** Enriquecer los metadatos de cada *chunk* con la 'fecha de acceso' o, si es posible, la 'fecha de publicación' del documento original.

3. **Citar la 'Antigüedad' de los Datos:** Modificar el *prompt* del LLM para que, en respuestas sensibles (ayudas, leyes), mencione la fecha de la última actualización de sus datos (ej. 'Según la información disponible a fecha de [FECHA DE INDEXADO],...').
4. **Sistema de Feedback:** Implementar en la interfaz de Streamlit botones de 'pulgar arriba / pulgar abajo' por cada respuesta. Este feedback permitiría identificar respuestas subóptimas y curar un 'dataset de grano fino' para futuras evaluaciones o afinamientos (fine-tuning).
5. **Disclaimer Dinámico:** Asegurar que el 'AVISO IMPORTANTE' mencionado en la documentación del proyecto esté siempre visible, reforzando que la herramienta es informativa y no sustituye la consulta experta.

5 Anexo: Tabla Resumen de Evaluación

Cuadro 1: Evaluación detallada de los casos de prueba.

Pregunta (Tema)	Pertinencia Chunks	Claridad Respuesta	Apoyo Práctico (Accionabilidad)
¿Trámites diagnóstico?	Alta	Alta	Alta (Proceso claro y contacto clave)
¿Ayudas económicas?	Alta	Alta	Muy Alta (Nombres de becas e importes)
¿Atención temprana Cádiz?	Alta	Alta	Alta (Explica el proceso, no el centro)
¿Derechos sistema educativo?	Alta	Alta	Alta (Lista derechos y equipos)
¿Centros en Sevilla?	Alta	Alta	Muy Alta (Menciona 'Autismo Sevilla')
¿Valoración discapacidad?	Muy Alta	Muy Alta	Muy Alta (Guía paso a paso)
¿Becas tratamientos?	Alta	Alta	Alta (Conecta beca NEAE con terapias)
¿Orientación educativa?	Muy Alta	Muy Alta	Muy Alta (Explica estructura EOEE)
¿Asociaciones de apoyo?	Alta	Alta	Alta (Identifica 'Autismo Andalucía')
¿Plazas educ. especial?	Muy Alta	Alta	Muy Alta (Define Modalidad 'D' y criterios)
¿Beneficios fiscales?	Media	Muy Alta	Muy Alta (Evita alucinación y reorienta)
¿Adaptaciones curriculares?	Alta	Alta	Muy Alta (Proceso claro: Orientación ->EOEE)
¿Renovar cert. discapacidad?	Muy Alta	Muy Alta	Muy Alta (Explica 'de oficio' y da teléfono)