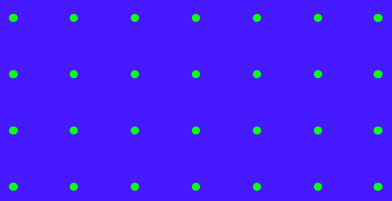




Uso de la

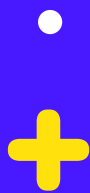
Generación Aumentada por Recuperación (RAG)

para mejorar la experiencia del cliente









ATENTO

ATENTO

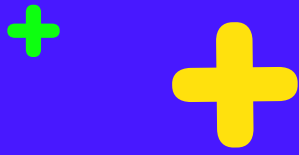


CONTENIDO



-  1 **Introducción**
-  2 **¿Qué es RAG?**
-  3 **Beneficios de RAG**
-  4 **Aplicaciones de RAG**
-  5 **Mecanismos integrados en la solución RAG de Atento**
-  6 **Conclusiones**





Introducción

En un mundo en el que la inteligencia artificial se ha convertido en la nueva normalidad, la **Generación Aumentada por Recuperación (RAG)** surge como una tecnología transformadora diseñada para **mejorar la precisión y la fiabilidad de los grandes modelos de lenguaje (LLM) mediante la incorporación de fuentes de conocimiento externas**. En Atento, hemos aprovechado la RAG para mejorar significativamente nuestro Knowledge Assistant, brindando a los usuarios respuestas precisas y contextualmente relevantes. Este White paper ofrece una exploración detallada de la RAG, sus ventajas, limitaciones, mejores prácticas y los mecanismos que hemos implementado para optimizar la experiencia del usuario.



¿Qué es la RAG?

La Generación Aumentada por Recuperación (Retrieval-Augmented Generation - RAG) es una tecnología híbrida de IA que combina la recuperación de información con la generación de texto para ofrecer respuestas precisas a las consultas de los usuarios. Este proceso se puede dividir en dos pasos principales:

STEP

01

Recuperación de información

El sistema busca documentos relevantes o fragmentos de texto de una amplia base de datos en respuesta a una consulta del usuario.

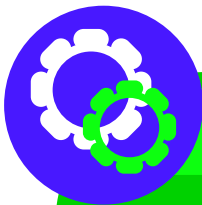


STEP

02

Generación de texto

A partir de los documentos recuperados, el sistema genera una respuesta informativa adaptada al contexto de la consulta.



Este enfoque permite una interacción dinámica, en la que el sistema **aprovecha la información existente para mejorar la calidad y la relevancia de las respuestas generadas.**





Beneficios de la RAG:



MAYOR PRECISIÓN Y RELEVANCIA

Al acceder a información actualizada, la RAG mejora la precisión y la relevancia de las respuestas.

ACCESO A LA INFORMACIÓN ACTUAL

La RAG proporciona acceso a la información más allá del límite de formación del LLM, lo que garantiza que las respuestas se basen en los datos más recientes.



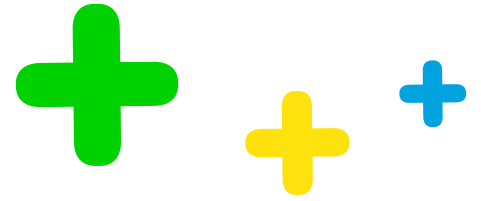
CREDIBILIDAD Y CONFIANZA

La capacidad de citar fuentes mejora la credibilidad y confiabilidad de la información proporcionada.

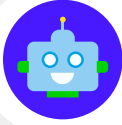
RENTABILIDAD

Mejorar las capacidades de LLM a través de RAG puede ser más rentable que reentrenar modelos con nuevos datos.





Aplicaciones de la RAG:



Chatbots e IA conversacional: Mejora el rendimiento de los chatbots al proporcionar respuestas más completas y contextuales.



Asistencia Profesional: Asiste a los profesionales con conocimientos especializados.



Atención al cliente: Mejora la calidad del soporte al proporcionar información precisa y relevante.

Limitaciones y mejores prácticas de las soluciones RAG

Limitaciones:

- 🟢 **Dependencia de datos:** La eficacia de RAG depende de la calidad y exhaustividad de la base de conocimientos externa.
- 🟢 **Complejidad de las preguntas:** Las preguntas complejas o ambiguas pueden poner a prueba la capacidad del sistema para recuperar información precisa.
- 🟢 **Costo computacional:** El proceso dual de recuperación y generación puede ser computacionalmente intensivo.
- 🟢 **Desafíos de la generación de texto:** Si los datos recuperados no están bien seleccionados, pueden ocurrir problemas como generar información incorrecta o "alucinaciones".

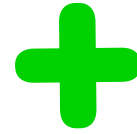




Mejores prácticas:

- **Preparación de documentos:** Mantén una base de datos actualizada y bien organizada con secciones jerárquicas y etiquetas relevantes.
- **Formulación de preguntas:** Formula preguntas de manera clara y específica para evitar ambigüedades y mejorar la precisión.
- **Mejora continua:** Cura y actualiza regularmente la base de datos para alinearla con las necesidades específicas del sistema RAG.





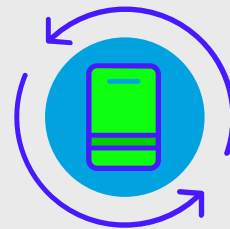
Mecanismos integrados en la solución RAG de Atento

Para mejorar el rendimiento y la experiencia de usuario de nuestro sistema RAG, Atento ha integrado varios mecanismos avanzados:



Interfaz fácil de usar

Interfaces que permiten a los usuarios controlar varios aspectos de la solución con una fricción mínima, incluido el cambio dinámico entre bases de conocimiento dentro de una sola conversación.



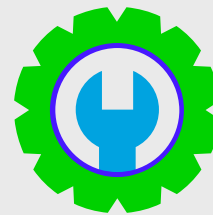
Enfoques de reclasificación

Después de la recuperación inicial, un paso opcional de reclasificación reorganiza los fragmentos de texto en función de criterios o metadatos adicionales, lo que mejora la relevancia.



Reformulación de preguntas

El sistema puede crear preguntas similares a la consulta del usuario, mejorando la precisión de las respuestas al considerar formulaciones de preguntas diversificadas.



Integración con los sistemas existentes

Integración perfecta con otros sistemas y canales de comunicación de la empresa, como Microsoft Teams y Hubbie, lo que mejora la flexibilidad operativa.



Conclusiones

La integración de la tecnología de **Generación Aumentada por Recuperación (RAG)** en los sistemas de IA ofrece un enfoque transformador para **mejorar la precisión, la relevancia y la fiabilidad de las respuestas generadas por los grandes modelos de lenguaje (LLM)**. Al aprovechar las fuentes de conocimiento externas, RAG aborda las limitaciones clave de los LLM tradicionales, como la información obsoleta o incorrecta.

La capacidad de la RAG para acceder a información actualizada garantiza que las respuestas se basen en los datos más actualizados disponibles, lo que **mejora la experiencia general del usuario**. Además, la incorporación de técnicas de reclasificación y reformulación de preguntas refina aún más la precisión y la relevancia contextual de las respuestas.

Sin embargo, la implementación exitosa de la RAG requiere una preparación meticulosa de documentos, una formulación de preguntas claras y una curación continua de la base de datos para mantener la eficacia del sistema. Comprender estas mejores prácticas es esencial para maximizar el potencial de RAG y garantizar su escalabilidad y eficiencia.



¿Estás list@ para pasar a la acción?

Contáctanos hoy mismo para saber cómo nuestras soluciones avanzadas de RAG pueden mejorar tus operaciones comerciales y ofrecer a tus clientes un servicio incomparable.



Atento.com

