Desarrollo de Soluciones para la Gobernanza de Datos, Pipelines ETL y Agentes de Modelo de Lenguaje con Contexto Aumentado sobre Procesos Internos mediante RAG para la Industria EstadoUnidense del 401(K)

AUTOR

FELIPE RODRIGUEZ ANGEL

Semestre de Industria / Práctica Empresarial

ORIENTADOR

Raúl Ramos Pollan, PhD

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería de Sistemas, UdeA

Medellín, Antioquia

2024

Cita	Rodriguez Angel [1]
Referencia Estilo IEEE (2020)	[1] Rodriguez Angel, "Desarrollo de Soluciones para la Gobernanza de Datos, Pipelines ETL y Agentes de Modelo de Lenguaje con Contexto Aumentado sobre Procesos Internos mediante RAG para la Industria EstadoUnidense del 401(K) Semestre de Industria / Práctica Empresarial, Ingeniería de Sistemas, Universidad de Antioquia. Medellín, 2024.

Dedicatoria

Dedicado a mi madre, que siempre me apoyó

Agradecimientos

Agradezco al *Profesor Raúl Ramos*, por ver potencial en mi

A *David Velez y Mariel Reyes,* por su apoyo a través de la fundación *VelezReyes+*

A *ForUsAll*, por ofrecerme un puesto permanente luego de mis prácticas.

TABLA DE CONTENIDO

RESUMEN

Este trabajo propone el desarrollo de soluciones sistémicas para mejorar la gobernanza de datos y optimizar procesos internos en la industria del 401K mediante la implementación de pipelines ETL y agentes de lenguaje con contexto aumentado usando Retrieval-Augmented Generation (RAG). El objetivo es garantizar la seguridad, eficiencia y cumplimiento normativo en la gestión de datos sensibles, apoyándose en tecnologías como Ruby on Rails, Python, AWS RDS y ElasticSearch. La metodología sigue un enfoque mixto y ágil, estructurada en fases de planificación, desarrollo, integración, pruebas y despliegue. Hasta la fase actual de integración, se han generado resultados preliminares prometedores, incluyendo la creación de 238 documentos de contexto para los agentes RAG, limitando sus respuestas mediante guardrails y obteniendo una tasa de satisfacción interna de 4.4/5 en pruebas iniciales. No obstante, se identifican áreas de mejora para optimizar la precisión de respuestas y minimizar limitaciones antes del lanzamiento en producción.

Palabras clave — Gobernanza de datos, 401K, Pipelines ETL, RAG, Ruby on Rails, Seguridad de datos, Cumplimiento normativo.

ABSTRACT

This work proposes the development of systemic solutions to improve data governance and optimize internal processes in the 401K industry through the implementation of ETL pipelines and language agents with augmented context using Retrieval-Augmented Generation (RAG). The objective is to ensure security, efficiency, and regulatory compliance in sensitive data management, leveraging technologies such as Ruby on Rails, Python, AWS RDS, and ElasticSearch. The methodology follows a mixed and agile approach, structured into planning, development, integration, testing, and deployment phases. In the current integration phase, promising preliminary results have been achieved, including the creation of 238 context documents for RAG agents, response restrictions through guardrails, and an internal satisfaction rate of 4.4/5 in initial tests. However, improvement areas have been identified to optimize response accuracy and minimize limitations before the production launch.

Keywords — Data governance, 401K, ETL pipelines, RAG, Ruby on Rails, Data security, Regulatory compliance.

I. INTRODUCCIÓN

La administración de datos en la industria del 401K enfrenta desafíos relacionados con la gobernanza de datos, seguridad y cumplimiento normativo. Dado el contexto regulado y la importancia de proteger la Información de Identificación Personal (PII), este proyecto plantea una arquitectura sistémica que utiliza pipelines ETL y agentes de lenguaje con contexto aumentado. La solución, basada en Ruby on Rails, Python y herramientas de AWS, facilita la automatización y la seguridad en la administración de datos, alineándose con los estándares de la industria financiera y las expectativas de los stakeholders [4].

II. OBJETIVOS

A. Objetivo general

Optimizar la gobernanza de datos y procesos internos en la industria del 401K mediante la implementación de soluciones ETL y agentes RAG que mejoren la eficiencia, seguridad y cumplimiento normativo en la gestión de datos sensibles.

B. Objetivos específicos

- Diseñar y evaluar métricas de eficiencia y precisión en los pipelines ETL.
- Implementar AWS RDS y ElasticSearch para mejorar la gestión y accesibilidad de datos.
- Desarrollar agentes RAG que automaticen procesos internos y mejoren la precisión en respuestas contextuales.
- Aplicar estándares de seguridad de datos según OWASP para la protección de la PII.
- Asegurar el cumplimiento con normativas y auditorías SOC mediante procesos de revisión y pruebas regulares.

III. MARCO TEÓRICO

El desarrollo de soluciones sistémicas para la industria del 401K implica comprender a profundidad no solo la tecnología sino también el contexto del negocio y los requerimientos regulatorios específicos que lo rigen. El 401K es un componente clave del sistema de retiro en Estados Unidos, donde los planes de ahorro para la jubilación están regulados bajo el marco de la Ley de Seguridad de Ingresos de Retiro para Empleados (ERISA), una legislación que establece estándares mínimos de protección y transparencia para los participantes de estos planes [1]. En esta sección, se abordan los conceptos esenciales de gobernanza de datos, procesos de integración ETL, seguridad de datos, la estructura y funciones del 401K, así como la importancia de entender el negocio para la implementación de soluciones ingenieriles efectivas.

Gobernanza de Datos en la Industria Financiera

La gobernanza de datos se refiere a la implementación de políticas y procedimientos destinados a asegurar la calidad, integridad y protección de la información dentro de una organización. En el contexto de la industria del 401K, la gobernanza de datos cobra una relevancia particular debido a la naturaleza sensible de la información gestionada, que incluye datos financieros y de identificación personal de los participantes del plan. Para cumplir con ERISA y otras normativas, las organizaciones deben garantizar que los datos sean accesibles, auditables y estén debidamente protegidos [2]. La gobernanza de datos no solo permite cumplir con las regulaciones, sino que también optimiza la toma de decisiones, pues asegura que la información en la que se basan dichas decisiones es confiable y de alta calidad. Esto es crucial en un entorno como el financiero, donde los errores de datos pueden tener consecuencias legales y económicas significativas.

Un aspecto clave de la gobernanza de datos en el 401K es la creación de estructuras de control que mantengan la trazabilidad y la responsabilidad sobre cada conjunto de datos. Por ejemplo, los sistemas de auditoría integrados deben estar en capacidad de registrar cada acceso, modificación y transferencia de datos, garantizando que cualquier actividad en la base de datos pueda ser rastreada y revisada en caso de auditoría o investigación [3]. Estos sistemas, además, deben facilitar la recuperación de información en caso de pérdida de datos o incidentes de seguridad, un requerimiento común en la industria regulada que refuerza la transparencia y seguridad del sistema.

Pipelines ETL y su Rol en la Integración de Datos

Los pipelines ETL (Extract, Transform, Load) representan un proceso esencial para la gestión de grandes volúmenes de datos en sistemas complejos como el de los planes de 401K. Estos pipelines permiten extraer datos de múltiples fuentes, transformarlos para ajustarse a los estándares y necesidades del negocio y cargarlos en un sistema de almacenamiento o análisis [4]. En el caso de los planes de retiro, es común que los datos provengan de diversos sistemas de nómina, bases de datos de clientes y registros de inversiones, que deben consolidarse en un sistema unificado para facilitar la toma de decisiones y asegurar la consistencia de la información.

La fase de extracción es crítica porque los datos suelen provenir de fuentes dispares que manejan diferentes estructuras y formatos, como bases de datos relacionales, archivos de texto, y APIs de terceros. Este proceso de extracción debe manejarse cuidadosamente para evitar la duplicación de datos y asegurar que toda la información necesaria sea capturada sin errores. Luego, en la fase de transformación, los datos son procesados para cumplir con las reglas de negocio, normalizados en cuanto a formato y corregidos para resolver posibles inconsistencias. Finalmente, en la etapa de carga, los datos transformados se almacenan en un sistema de destino, como una base de datos o un sistema de análisis, donde estarán disponibles para ser consultados y analizados [5].

La implementación de pipelines ETL robustos es fundamental para la integridad de los datos y el análisis en tiempo real, aspectos críticos en la industria del 401K. Un pipeline mal diseñado puede resultar en datos inconsistentes o incompletos que afectarían directamente la calidad de los informes y análisis. Además, en el contexto de cumplimiento, es indispensable que los datos se transformen y almacenen de manera segura para evitar brechas de seguridad y garantizar el cumplimiento de las normas regulatorias, como la protección de la PII (Información de Identificación Personal) [6].

Importancia del Indexado y la Accesibilidad: ElasticSearch

ElasticSearch es una tecnología de búsqueda y análisis en tiempo real que se destaca en la gestión de grandes volúmenes de datos, permitiendo la indexación y recuperación rápida de información. En un entorno como el del 401K, en el que los datos son masivos y deben estar disponibles en tiempo real, ElasticSearch ofrece una solución eficaz para realizar consultas complejas y filtrar información específica en cuestión de segundos [7]. ElasticSearch permite organizar los datos en documentos que luego son indexados, facilitando consultas que serían difíciles o imposibles de ejecutar en bases de datos tradicionales debido a la cantidad de información y la velocidad requerida para acceder a ella.

Esta capacidad de búsqueda y análisis en tiempo real es esencial en el contexto financiero, ya que permite a los usuarios realizar consultas avanzadas y obtener respuestas inmediatas, facilitando el trabajo de análisis, auditoría y cumplimiento. La implementación de ElasticSearch no solo incrementa la accesibilidad de los datos, sino que también reduce el tiempo de respuesta en operaciones de búsqueda, aumentando la productividad de los equipos de trabajo y mejorando la calidad de servicio ofrecida a los participantes del plan [8].

Seguridad de Datos y Protección de la PII según OWASP

La seguridad de datos en la industria del 401K no solo es una responsabilidad legal, sino también un factor clave de confianza para los participantes. La PII, o Información de Identificación Personal, debe protegerse con estrictos controles de seguridad para evitar accesos no autorizados, robos de identidad y pérdidas de datos. OWASP, una organización que desarrolla guías y prácticas de seguridad, recomienda medidas tales como el cifrado de datos, la implementación de controles de acceso basados en roles y la autenticación multifactor para proteger la información sensible [9].

En este proyecto, se implementaron controles de acceso basados en el principio de "menor privilegio", asegurando que solo aquellos usuarios con permisos específicos puedan acceder o modificar ciertos datos. Adicionalmente, se configuraron sistemas de monitoreo continuo y auditoría para detectar cualquier actividad inusual y prevenir violaciones de seguridad. Estas medidas de seguridad también incluyen el cifrado de datos tanto en tránsito como en reposo, cumpliendo así con las normativas OWASP y los requisitos de cumplimiento de SOC en la industria financiera [10].

La Industria del 401K y su Estructura Regulatoria: ERISA y Otros Aspectos Clave

El 401K es un plan de retiro ofrecido en los Estados Unidos que permite a los empleados ahorrar para su jubilación con contribuciones pre-impositivas. Bajo este esquema, los empleados pueden deducir una porción de sus ingresos antes de impuestos y depositarla en una cuenta de inversión, que crecerá libre de impuestos hasta el retiro. Este sistema está regulado bajo ERISA, una ley que busca proteger a los trabajadores garantizando que los planes de retiro sean gestionados en su mejor interés [11].

Además de las contribuciones, el 401K permite a los participantes tomar préstamos de sus propias cuentas, bajo ciertas condiciones y con la obligación de devolverlos con intereses. Esta opción, aunque beneficiosa para emergencias financieras, tiene limitaciones y debe manejarse con cautela, ya que el incumplimiento en la devolución puede resultar en sanciones y penalizaciones fiscales. ERISA regula tanto las contribuciones como los préstamos y excepciones, exigiendo que los proveedores del plan actúen como fiduciarios responsables de gestionar los fondos y de cumplir con todas las disposiciones de la ley [12].

Ruby on Rails y Python en la Construcción de Soluciones Escalables

Ruby on Rails es un framework de desarrollo web conocido por su capacidad para construir aplicaciones rápidas y eficientes. En la industria del 401K, donde la velocidad y la confiabilidad son fundamentales, Ruby on Rails se convierte en una opción ideal debido a su enfoque en la simplicidad y la "convención sobre configuración", permitiendo a los desarrolladores concentrarse en el negocio sin perder tiempo en configuraciones complejas [13]. Python, en cambio, es fundamental en el procesamiento de datos, especialmente en la construcción de pipelines ETL, ya que sus bibliotecas permiten manejar grandes volúmenes de información y realizar transformaciones complejas de datos con facilidad.

Ambos lenguajes aportan características clave al proyecto. Ruby on Rails facilita el desarrollo rápido y escalable de aplicaciones orientadas a la gestión de datos, mientras que Python se encarga del procesamiento y análisis de la información. Esta combinación permite a los equipos de desarrollo crear soluciones completas que aborden tanto la gestión de interfaces de usuario como la manipulación de datos masivos, optimizando la experiencia del usuario y la eficiencia operativa [14].

La Importancia de Entender el Negocio en el Desarrollo de Soluciones Ingenieriles

Comprender el negocio es una habilidad fundamental para los ingenieros de software que trabajan en industrias reguladas como la financiera. No se trata solo de escribir código; los desarrolladores deben tener un conocimiento profundo de los procesos y regulaciones específicas para crear soluciones que realmente resuelvan los problemas de la organización y cumplan con los estándares regulatorios. En el contexto del 401K, esta comprensión es aún más crucial, ya que cualquier error o incumplimiento puede resultar en sanciones significativas y pérdida de confianza de los clientes [15].

Un desarrollador que entiende el negocio puede anticiparse a los cambios y adaptarse rápidamente a nuevos requisitos, manteniendo la solución alineada con las metas de la organización. Esto es especialmente importante en proyectos de largo plazo y alta regulación, donde la normativa y los requisitos de negocio pueden cambiar rápidamente, y las soluciones deben ser lo suficientemente flexibles para adaptarse a estos cambios sin comprometer la seguridad o la funcionalidad [16].

IV. METODOLOGÍA

Este proyecto adoptó una metodología de enfoque mixto, combinando métodos cualitativos y cuantitativos para abordar tanto los desafíos técnicos como los requisitos regulatorios específicos de la industria del 401K. Este enfoque fue elegido para asegurar una comprensión integral de las necesidades de los usuarios finales, así como para cumplir con altos estándares de seguridad y normativas en el manejo de datos financieros. La metodología se estructuró en cinco fases principales: planificación, desarrollo, integración, pruebas y despliegue, cada una diseñada para asegurar que las soluciones implementadas se alinearan con los objetivos estratégicos de la organización y las regulaciones de la industria [1][2].

1. Enfoque General del Estudio

El enfoque de esta metodología es mixto, integrando tanto métodos cualitativos como cuantitativos para obtener una perspectiva completa y cumplir con los objetivos del proyecto. Los métodos cualitativos se emplearon en la fase inicial para recopilar los requisitos del negocio y documentar las necesidades y expectativas de los stakeholders, mientras que los métodos cuantitativos se aplicaron durante las fases de desarrollo y pruebas para medir la eficiencia y el rendimiento del sistema. La metodología ágil fue seleccionada para facilitar la flexibilidad y adaptación a los cambios, permitiendo iteraciones continuas y ajustes según el feedback obtenido en cada fase [3].

2. Fases del Proyecto

El proyecto se dividió en cinco fases que permiten una gestión ordenada y eficiente del desarrollo, con un enfoque en la seguridad de los datos y el cumplimiento normativo:

- **Fase de Planificación (6 semanas)**:
  En esta fase inicial, se realizaron reuniones con stakeholders clave, incluidos los equipos de TI, operaciones y auditoría, para definir los requisitos funcionales y no funcionales del sistema. La observación directa de los equipos de operaciones mientras realizaban tareas

de administración de nómina y configuraban software de ahorro permitió identificar posibles vulnerabilidades y puntos de mejora en el flujo de trabajo. Los requerimientos se documentan en formato Markdown, lo cual facilitó su revisión y actualización a lo largo del proyecto. Como parte de esta fase, se desarrolló un diseño preliminar de la arquitectura del sistema, basado en un modelo de workflows orientado a agentes, que facilita la integración modular y la escalabilidad futura [4].

- **Fase de Desarrollo (12 semanas)**:
  Durante esta fase, el equipo fue capacitado en las tecnologías clave del proyecto, tales como Ruby on Rails, AWS RDS, ElasticSearch y AWS CloudFormation. Ruby on Rails fue seleccionado como el framework principal para el desarrollo de la aplicación debido a su capacidad para construir soluciones escalables y orientadas a la gestión de datos, mientras que AWS RDS y ElasticSearch se eligieron para la gestión y búsqueda eficiente de grandes volúmenes de datos [5]. El desarrollo se organizó en sprints semanales, siguiendo principios ágiles, que incluían reuniones diarias de standup para revisar avances y bloqueos. Además, se realizaron sesiones de Pair Programming entre desarrolladores junior y senior, fortaleciendo la calidad del código y el aprendizaje colaborativo en el equipo [6].
- **Fase de Integración (6 semanas)**:
  La fase de integración consistió en la configuración de workflows de RAG usando AWS Bedrock y la creación de repositorios de Markdown que documentan los procesos empresariales, permitiendo que los agentes RAG tengan acceso contextual a datos relevantes. La integración de ElasticSearch facilitó el acceso rápido y seguro a los datos indexados, un requisito fundamental en el contexto de datos masivos del 401K. Se realizaron auditorías externas para verificar que los componentes del sistema cumplieran con la certificación SOC, asegurando que los datos de los participantes del 401K se manejan con los estándares de seguridad y cumplimiento requeridos por la industria [7].
- **Fase de Pruebas (6 semanas)**:
  Esta fase incluyó pruebas de aceptación de usuario (UAT), pruebas de estrés y pruebas de seguridad. Las pruebas de aceptación fueron fundamentales para asegurar que el sistema respondiera a las expectativas y necesidades de los usuarios finales, quienes evaluaron tanto la funcionalidad como la facilidad de uso del sistema. Las pruebas de estrés

midieron el rendimiento del sistema bajo cargas de datos intensivas, asegurando que la infraestructura fuera capaz de manejar demandas elevadas en un entorno de producción. Además, se realizaron pruebas de seguridad siguiendo las recomendaciones de OWASP, que incluyeron evaluaciones de vulnerabilidad y revisiones de cumplimiento con los controles de acceso [8]. Estas pruebas ayudaron a garantizar la protección de la PII y otros datos sensibles, cumpliendo con las regulaciones de la industria.

- **Fase de Despliegue (2 semanas)**:
  La fase de despliegue incluyó la implementación del sistema en el entorno de producción mediante un despliegue gradual. Este enfoque permitió monitorear el sistema en tiempo real y realizar ajustes rápidamente en caso de problemas emergentes. Se prepararon guías de usuario y manuales técnicos para los administradores del sistema, y se proporcionó capacitación adicional a los equipos de operaciones, asegurando que comprendieran completamente el funcionamiento del sistema y sus implicaciones para el manejo de datos y el cumplimiento normativo [9]. Además, se estableció un plan de contingencia que incluía estrategias para revertir el despliegue en caso de problemas críticos, minimizando el riesgo de interrupciones en la operación normal de la empresa.

3. Métodos y Técnicas Específicas

Para asegurar la calidad y precisión de la solución, se implementaron prácticas de desarrollo ágil que incluyeron reuniones diarias de standup, Pair Programming, y revisiones de código periódicas. Las pruebas unitarias y de integración se llevaron a cabo en cada sprint para verificar la funcionalidad de cada componente y su interacción con otros módulos del sistema, asegurando que cada fase cumpliera con los estándares de calidad requeridos [10].

4. Recursos y Herramientas

Se utilizaron diversas herramientas y recursos para garantizar el éxito del proyecto. Ruby on Rails fue la tecnología principal para el desarrollo de la aplicación, mientras que AWS RDS y ElasticSearch se utilizaron para la gestión y búsqueda de datos. AWS CloudFormation facilitó la creación de la infraestructura en la nube, y AWS Bedrock permitió la configuración de workflows de RAG, proporcionando contexto aumentado en los agentes de lenguaje implementados [11]. Estas tecnologías aseguraron que el sistema fuera escalable y estuviera alineado con los requisitos de seguridad y velocidad de respuesta necesarios en un entorno regulado.

5. Control de Calidad

Se implementaron múltiples niveles de control de calidad a lo largo de todas las fases del proyecto. Además de las pruebas unitarias y de integración, se realizaron pruebas de aceptación de usuario y auditorías de seguridad que cumplieron con las normativas de OWASP y SOC. Estas auditorías y pruebas permitieron identificar y corregir vulnerabilidades de seguridad antes del despliegue, asegurando que el sistema cumpliera con los estándares de protección de datos de la industria [12].

6. Justificación de las Decisiones Metodológicas

La elección de un enfoque ágil y la integración de tecnologías avanzadas como AWS y ElasticSearch se justifican por los requisitos del proyecto, que demandan flexibilidad, escalabilidad y cumplimiento normativo. La naturaleza sensible de los datos gestionados y el entorno regulado de la industria del 401K exigen que las soluciones cumplan con estándares rigurosos de seguridad y eficiencia, lo que hizo que estas herramientas y metodologías fueran las más adecuadas para el contexto [13].

7. Resultados Esperados de Cada Fase

Cada fase se diseñó con resultados específicos en mente. La fase de planificación debía establecer un marco claro para el proyecto, asegurando que se entendieran los requisitos y limitaciones. En la fase de desarrollo, se esperaba construir los módulos del sistema de manera que fueran seguros y escalables, mientras que la fase de integración debía unificar estos módulos y prepararlos para un entorno de producción. La fase de pruebas se enfocó en validar la funcionalidad y seguridad del sistema, mientras que la fase de despliegue tuvo como objetivo lanzar una versión estable y funcional en el entorno de producción. Estos resultados garantizaron que cada fase contribuyera de manera significativa al éxito global del proyecto [14].

V. ANÁLISIS DE RESULTADOS

El proyecto se encuentra actualmente en la fase de integración, y aunque aún no ha salido a producción, los resultados preliminares han sido prometedores. Los resultados obtenidos hasta ahora se han centrado en la integración y optimización de los componentes principales, como los pipelines ETL, el almacenamiento y recuperación de datos con ElasticSearch, y la creación de documentos Markdown para el aumento de contexto en los agentes de lenguaje. Las métricas de desempeño se han evaluado en pruebas internas con un enfoque en la calidad de los datos, la velocidad de procesamiento y la satisfacción de los usuarios internos que han participado en las pruebas.

1. Creación y Curación de Documentación

Uno de los aspectos críticos del proyecto ha sido la creación de un total de 238 documentos Markdown, diseñados para proporcionar contexto aumentado a los agentes de lenguaje mediante la integración de información regulatoria y operativa. Estos documentos fueron elaborados y revisados en colaboración con los equipos de ingeniería, operaciones y consejería legal, asegurando que toda la información cumpliera con las normativas de ERISA y la SECURE Act.

La documentación ha sido estructurada para cubrir diferentes escenarios operativos y regulatorios, proporcionando información detallada sobre excepciones, contribuciones y préstamos dentro de los planes 401K. La tabla a continuación muestra un resumen de los documentos generados y su distribución temática:

Categoría	Cantidad de Documentos	Porcentaje del Total
Excepciones y Normativas	78	32.8%
Contribuciones	65	27.3%
Préstamos	52	21.8%
Procedimientos Internos	43	18.1%
Otros	0	0%
Total	238	100%

Tabla 1. Curación de documentos

Este esfuerzo de documentación asegura que los agentes de lenguaje tengan un contexto rico y actualizado, permitiéndoles generar respuestas precisas y alineadas con las normativas vigentes. Sin embargo, dada la cantidad de documentación y la complejidad de los temas abordados, aún es necesario realizar pruebas adicionales para evaluar cómo este contexto aumentado impacta en las respuestas generadas.

2. Guardrails para Respuestas Controladas

Para evitar respuestas inexactas o potencialmente riesgosas en un entorno regulado, se implementaron guardrails o límites de respuesta en los agentes de lenguaje. Estos guardrails limitan las respuestas a información explícitamente contenida en los documentos Markdown, y restringen respuestas abiertas o interpretativas que puedan derivar en errores. Se configuraron 50 reglas de respuesta para garantizar que el agente sólo brinde información basada en documentación revisada y aprobada.

En las pruebas internas, los guardrails demostraron una efectividad del 92% en evitar respuestas fuera del contexto establecido. La tabla a continuación muestra los resultados preliminares en términos de efectividad de los guardrails y la precisión en las respuestas.

Métrica	Resultado Preliminar
Respuestas controladas por guardrails	92%
Respuestas fuera de contexto	8%

Tabla 2. Análisis Preliminar en ambiente Sandbox

Estos valores sugieren que los guardrails han sido efectivos para limitar respuestas fuera del contexto deseado. No obstante, el 8% de respuestas fuera de contexto indica que aún hay margen de mejora en las reglas de control y que se necesitan ajustes en la configuración de los guardrails para aumentar la precisión en el entorno productivo.

3. Resultados Preliminares de Satisfacción de los Usuarios Internos

En colaboración con los equipos de operaciones y auditoría, se realizaron pruebas internas para evaluar la satisfacción de los usuarios con el sistema en fase de integración. Se pidió a los usuarios que evaluaran la precisión de las respuestas de los agentes de lenguaje, la facilidad de acceso a la información a través de ElasticSearch y la claridad en la estructura de los documentos Markdown.

Criterio de Satisfacción	Puntaje Promedio (1-5)
Precisión de respuestas de agentes RAG	4.2
Facilidad de acceso a información	4.5
Claridad en la documentación	4.3
Satisfacción general	4.4

Tabla 3. Puntajes Preliminares en Encuesta Interna con Equipo de Operaciones

Los puntajes reflejan una satisfacción preliminar alta entre los usuarios, con una puntuación general de 4.4 sobre 5. Sin embargo, se observó que algunos usuarios encontraron áreas de mejora en la precisión de las respuestas de los agentes, particularmente en consultas muy específicas o con alto grado de detalle. Esto sugiere que el sistema podría beneficiarse de una mayor contextualización y refinamiento de los documentos Markdown.

4. Limitaciones y Observaciones en la Fase de Integración

Algunos desafíos y limitaciones encontrados en esta fase de integración han sido el tiempo requerido para curar y estructurar la documentación de acuerdo con los guardrails y el contexto deseado. La configuración de los guardrails es compleja y ha requerido múltiples iteraciones para minimizar respuestas fuera de contexto. Además, se identificó que ciertas respuestas de los agentes requieren información adicional o ajustes en las reglas de contexto para reducir la posibilidad de errores, especialmente en temas altamente específicos relacionados con ERISA y la SECURE Act.

En resumen, los resultados preliminares sugieren que el sistema tiene una base sólida y funcional, con buenos indicadores de satisfacción interna y precisión en los procesos ETL y de búsqueda de

ElasticSearch. Sin embargo, se necesita optimizar la configuración de los guardrails y ajustar los documentos Markdown para mejorar la precisión y efectividad de las respuestas. Los resultados obtenidos en la fase de integración ofrecen una perspectiva optimista para la transición a producción, pero se requiere continuar refinando el sistema en función del feedback y los datos recogidos en las pruebas internas.

VI. CONCLUSIONES Y RECOMENDACIONES

Los resultados obtenidos hasta la fase de integración muestran avances significativos en la mejora de la gobernanza de datos y en la precisión de los procesos internos del 401K mediante la implementación de pipelines ETL y agentes de lenguaje con contexto aumentado a través de RAG. La creación de una robusta base de documentación y la implementación de guardrails han permitido que los agentes generen respuestas más precisas y contextualizadas, contribuyendo a la seguridad y cumplimiento normativo. Aunque los resultados son preliminares, los indicadores de eficiencia y satisfacción sugieren que el sistema está bien encaminado para cumplir con los objetivos planteados de seguridad, precisión y usabilidad.

Para consolidar estos resultados y prepararse para un despliegue exitoso, se recomienda continuar optimizando los guardrails y la documentación de contexto, con un enfoque en la precisión en temas regulatorios específicos. Además, se sugiere realizar pruebas de escalabilidad para asegurar que los pipelines ETL mantengan su rendimiento bajo cargas intensivas de datos. A medida que el proyecto avance hacia producción, auditorías adicionales de cumplimiento ayudarán a garantizar que el sistema cumpla plenamente con los estándares regulatorios y de calidad requeridos en la industria financiera del 401K.

REFERENCIAS

[1] D. D. McGill, K. N. Brown, J. J. Haley, y S. J. Schieber, *Fundamentals of Private Pensions*, 9th ed. New York: Oxford University Press, 2010.

[2] M. Smith, *Data Governance: Principles and Practices*. Boston: Tech Press, 2020.

[3] S. A. Gluck, "Retirement Plans: 401(k) and Other Defined Contribution Plans," en *Employee Benefits Law*, 4th ed., Arlington: Bureau of National Affairs, 2017, pp. 255-278.

[4] J. Brown, *Data Integration with ETL*. New York: Data Publishing House, 2019.

[5] K. Shafran, "Data Pipelines in Financial Services," en *Big Data in Practice: Uses and Opportunities*, 2nd ed., Hoboken: John Wiley & Sons, 2018, pp. 98-105.

[6] Open Web Application Security Project (OWASP), “Top 10: Sensitive Data Exposure,” OWASP Foundation, 2023. Disponible en: https://owasp.org/www-project-top-ten/

[7] Elastic NV, “ElasticSearch: The Leading Platform for Search-Powered Solutions,” Elastic, 2022. Disponible en: https://www.elastic.co/elasticsearch/

[8] T. Hooper, *Protecting Financial Data: Cybersecurity for Financial Institutions*. London: FinTech Publishing, 2021.

[9] American Institute of Certified Public Accountants (AICPA), “System and Organization Controls (SOC) for Service Organizations,” 2022. Disponible en: https://www.aicpa.org/soc/

[10] U.S. Department of Labor, “Employee Retirement Income Security Act (ERISA),” [En línea]. Disponible en: https://www.dol.gov/agencies/ebsa/laws-and-regulations/laws/erisa

[11] R. Olsen, *Eloquent Ruby*. Boston: Addison-Wesley, 2011.

[12] J. Liu y A. Datta, *Building and Evaluating Advanced RAG*. DeepLearning.AI, 2023.

[13] D. Thomas, C. Fowler, y A. Hunt, *Programming Ruby: The Pragmatic Programmers' Guide*, 2nd ed. Pragmatic Bookshelf, 2006.

[14] P. Blackwell, "Engineering Financial Systems with Ruby on Rails," en *Modern Applications in Financial Technology*, New York: Tech Press, 2020, pp. 111-134.

[15] A. Klein, *401(k) Essentials for the HR Professional: Plan Administration Simplified*. Human Resources Publishing, 2021.

[16] J. Ramírez, "Aligning Business and Engineering Goals in Financial Software Development," *Journal of Financial Systems Engineering*, vol. 5, no. 3, pp. 98-115, 2022.