



COLEGIO DE CIENCIAS E INGENIERÍAS

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

Planificación para el Desarrollo del Proyecto Integrador

Tutor: Daniel Riofrío

Autor: Pablo Andrés Encalada Andrade

Quito – Ecuador
2025



1. Título del Proyecto:

Modelo de Recuperación Aumentada por Generación (RAG) para la Generación Automática de Pliegos en Compras Públicas de Subasta Inversa Electrónica en Ecuador.

2. Relevancia y Justificación:

La generación de pliegos en los procesos de compras públicas es una tarea compleja y demandante de tiempo. En el caso de Ecuador, las subastas inversas electrónicas representan un tipo de compra pública de alto uso pero que deben cumplir estrictamente con la normativa vigente, integrando cláusulas legales, requisitos técnicos y parámetros administrativos en un marco regulatorio que cambia con frecuencia. Este contexto genera una alta carga de trabajo para los equipos responsables y una curva de aprendizaje considerable para nuevos analistas

Elaborar pliegos de forma manual no solo implica un esfuerzo significativo de lectura y redacción, sino que también aumenta el riesgo de errores y omisiones que afectan la calidad y la validez de los procesos de contratación pública. Por ello, resulta necesario un enfoque que permita optimizar la generación documental, estandarizando su calidad y asegurando consistencia normativa.

Bajo este contexto, los modelos Retrieval-Augmented Generation (RAG) ofrecen una alternativa tecnológica para automatizar la generación de pliegos, al combinar mecanismos de recuperación de normativa y datos relevantes con módulos generativos capaces de redactar textos humanizados y alineados a la ley vigente. Este enfoque no solo asegura que los pliegos generados estén anclados en evidencia normativa, sino que además permite integrar procesos de validación automáticos y manuales para garantizar su corrección y trazabilidad.

En particular, el presente proyecto utilizará como base los archivos .ushay de Subasta Inversa Electrónica en Ecuador, junto con la normativa aplicable, para conformar un corpus documental estructurado. A partir de un prompt con los detalles de la contratación, el sistema RAG generará un pliego inicial conforme a la ley ecuatoriana, el cual será sometido a validaciones automáticas (consistencia, completitud, referencias normativas) y validaciones manuales mediante una interfaz gráfica de usuario.

Como contribución práctica, el presente proyecto propone la construcción de un sistema de Recuperación Aumentada por Generación (RAG) que automatiza la elaboración de pliegos para Subasta Inversa Electrónica en Ecuador, utilizando archivos .ushay y la normativa vigente como base de conocimiento. Además de la incorporación de una interfaz gráfica de usuario que permite a los analistas revisar, validar manualmente los pliegos generados, garantizando así el cumplimiento legal y reduciendo el riesgo de errores. Este esquema de validación dual, automática para consistencia y completitud, y manual a través de la interfaz, asegura que el sistema no solo acelere la generación de documentos, sino que también preserve la trazabilidad normativa y el control experto sobre los resultados. En base a esto, se anticipa una mejora sustantiva en precisión, cobertura y tiempo de respuesta, así como una curva de aprendizaje más corta para nuevos usuarios.



3. Objetivos

a. Generales

El objetivo de presente proyecto es desarrollar un modelo de Recuperación Aumentada por Generación (RAG) capaz de generar automáticamente pliegos para procesos de Subasta Inversa Electrónica en Ecuador, utilizando como corpus normativo la legislación vigente y como base de datos los archivos .ushay, garantizando trazabilidad normativa y facilitando validaciones automáticas y manuales.

b. Específicos

- Construir y/o curar el corpus documental a partir de normativa ecuatoriana vigente y los archivos .ushay de Subasta Inversa Electrónica, procesando sus estructuras XML para normalizar la información.
- Diseñar la arquitectura RAG que combine estrategias de recuperación híbrida (léxica y vectorial), embeddings adaptados al dominio legal/administrativo en español y mecanismos de re-ranking para priorizar evidencia normativa relevante.
- Implementar una capa de generación con guardrails de fidelidad, empleando plantillas de prompting orientadas a citación automática, control de alucinaciones y políticas de “responder solo con base en el contexto”.
- Construir una interfaz gráfica de usuario (GUI) que facilite la interacción con el sistema, incluyendo:
 - Generación inicial de pliegos.
 - Visualización de las referencias normativas recuperadas.
 - Validación manual complementaria a las verificaciones automáticas.
- Evaluar el desempeño del sistema mediante métricas de recuperación (recall, precisión, nDCG) y de generación (fidelidad normativa, relevancia, utilidad práctica), integrando herramientas como OPIK para la evaluación automatizada y comparando resultados frente al flujo manual tradicional.
- Documentar el sistema y sus resultados, incluyendo la arquitectura técnica, corpus procesado, decisiones de diseño, manual de uso de la interfaz y análisis de errores, estableciendo recomendaciones y líneas de trabajo futuro.



4. Estado del Arte

La contratación pública concentra una fracción significativa del gasto estatal y, por su complejidad, ha sido objeto de múltiples estudios en ciencia de datos. Si bien la literatura “data-driven” ha crecido, muchas revisiones señalan que se ha privilegiado el análisis de datos estructurados (montos, adjudicaciones, plazos) y se han infrautilizado los textos de pliegos, contratos y resoluciones, que constituyen insumos críticos para garantizar consistencia y validez normativa [1]. En este contexto, los modelos de Retrieval-Augmented Generation (RAG) representan una línea emergente que combina recuperación de evidencia con generación condicionada, reduciendo alucinaciones y habilitando respuestas citables, un enfoque especialmente pertinente en dominios regulados como el jurídico-administrativo.

En la literatura más cercana al procesamiento automatizado de pliegos, destaca el trabajo de Katona y Fazekas [2], un proyecto de minería de texto que inspecciona directamente el contenido de pliegos para identificar cláusulas restrictivas y barreras a la competencia. Aunque su énfasis está en la detección de prácticas opacas, lo relevante metodológicamente es el uso directo de documentos textuales como fuente de evidencia.. De manera similar, la ponencia presentada en el ECPR (2023) confirma que los enfoques basados en texto permiten recuperar y citar cláusulas específicas en pliegos y contratos, lo que respalda la pertinencia de metodologías centradas en evidencia documental [1].

En la misma línea, los indicadores de riesgo y red flags validados con datos nacionales, como el estudio en EPJ Data Science sobre obras viales en Italia y la propuesta de indicadores multilayer en los Quaderni de la UIF-Banca d’Italia, muestran que esto es algo alcanzable e incluso escalable, pero carecen de un componente de generación ni validación documental [3, 4], precisamente el vacío que el presente proyecto busca cubrir mediante automatización de documentos normativos.

En el ámbito de plataformas, experiencias como ProZorro y DOZORRO en Ucrania muestran cómo el uso de inteligencia artificial y machine learning puede priorizar licitaciones de alto riesgo y canalizar reportes ciudadanos [5, 4]. No obstante, su enfoque es principalmente analítico y de vigilancia, sin un componente de generación automática de pliegos o documentos base.

Finalmente, en el área legal, se desarrolló un sistema llamado LegalBench-RAG, un benchmark que evalúa la calidad de recuperación en contextos jurídicos y demuestra que la fidelidad de las respuestas depende en gran medida de la calidad del retriever y de los mecanismos de citación [6]. Este hallazgo es directamente transferible a la contratación pública, puesto que, para que un sistema de generación de pliegos sea confiable, debe estar respaldado por recuperación robusta y citación precisa.

En definitiva, la automatización de documentos normativos en compras públicas requiere integrar y consultar grandes volúmenes de información (normativas, pliegos, contratos, etc.). Los enfoques plenamente generativos (LLMs) ofrecen accesibilidad en lenguaje natural, pero alucinan y carecen de citabilidad si no están anclados en evidencia. Es por esto por lo que surge la necesidad de Retrieval-Augmented Generation (RAG), que combina recuperación desde un corpus curado con generación condicionada por los documentos recuperados, priorizando fidelidad y trazabilidad.

A continuación, se describen conceptos relevantes al área de estudio:

a) Contratación pública y pliegos en Ecuador



La contratación pública es el conjunto de procedimientos mediante los cuales el Estado adquiere bienes, servicios u obras con el fin de atender las necesidades colectivas. Está regida por principios de legalidad, transparencia, igualdad, libre concurrencia y eficiencia en el uso de recursos públicos, garantizando que las adquisiciones se realicen bajo reglas claras y con trazabilidad [7].

En este ámbito, el Estado actúa como contratante y las personas naturales o jurídicas como oferentes. La Ley Orgánica del Sistema Nacional de Contratación Pública (LOSNCOP), publicada en el Registro Oficial Suplemento 395 del 4 de agosto de 2008 y reformada en varias ocasiones, establece los procedimientos, modalidades y requisitos que deben observarse. La administración de este sistema corresponde al Servicio Nacional de Contratación Pública (SERCOP), que emite resoluciones, manuales y guías obligatorias para las entidades contratantes [8].

La contratación pública tiene gran relevancia porque concentra una parte significativa del gasto estatal y constituye una herramienta clave para garantizar el acceso equitativo de proveedores a los procesos de compra. En el caso ecuatoriano, el Sistema Nacional de Contratación Pública (SNCP) ofrece modalidades como licitación, cotización, subasta inversa electrónica, catálogo electrónico y régimen especial, entre otras [9].

b) Subasta Inversa Electrónica (SIE)

La Subasta Inversa Electrónica es un procedimiento de contratación pública utilizado principalmente para la adquisición de bienes y servicios estandarizados. Se caracteriza porque los proveedores compiten ofreciendo precios a la baja en un proceso dinámico y en línea, lo que fomenta la eficiencia y la transparencia [8]. En este procedimiento, existe una entidad contratante que fija un precio referencial y condiciones iniciales mediante un pliego. Los oferentes realizan sus propuestas en rondas sucesivas, reduciendo sus precios hasta alcanzar la mejor oferta dentro de los parámetros establecidos.

En Ecuador, la SIE está regulada por la Ley Orgánica del Sistema Nacional de Contratación Pública (LOSNCOP) y por las resoluciones emitidas por el SERCOP, que definen su operatividad y condiciones obligatorias para los pliegos. Cada proceso queda registrado en archivos .ushay en formato XML, que almacenan la información estructurada del procedimiento.

c) Pliegos

Un pliego es el documento oficial que regula un proceso de contratación pública. Este documento contiene las condiciones legales, técnicas y administrativas que deben cumplirse para llevar a cabo una contratación y es obligatorio tanto para la entidad contratante como para los oferentes. Dentro del pliego, existen referencias normativas, requisitos de habilitación de los proveedores, especificaciones técnicas del bien o servicio requerido, condiciones contractuales como plazos y garantías, criterios de evaluación de ofertas y, en muchos casos, un modelo de contrato [7].

En Ecuador, la elaboración de pliegos está regulada por la Ley Orgánica del Sistema Nacional de Contratación Pública (LOSNCOP) y las normas emitidas por el SERCOP. En la Subasta Inversa Electrónica (SIE), los pliegos son especialmente importantes porque definen



el precio referencial, los ítems a subastar, las reglas de participación y las condiciones de adjudicación [8]. Su importancia radica en que garantizan la legalidad del procedimiento, aseguran transparencia y permiten comparar objetivamente las ofertas. Por eso, son la base documental crítica de cada proceso de contratación y un punto estratégico para aplicar soluciones tecnológicas de automatización como el modelo RAG.

Las subastas inversas electrónicas se ejecutan íntegramente en la plataforma informática del SERCOP. Cada uno de estos procesos generan un archivo en formato .ushay, que contiene estructuras en XML con los datos de configuración, requisitos y resultados del proceso. Estos archivos conforman la base documental y técnica de las subastas, y son utilizados para auditoría, análisis y control posterior [8]. Para este proyecto, los archivos .ushay correspondientes al período 2020–2025 serán la fuente de datos estructurada oficial, complementada con la normativa vigente, lo que permitirá construir un corpus normativo-documental sobre el cual aplicar técnicas de Recuperación Aumentada por Generación.

d) Fundamentos de Retrieval-Augmented Generation (RAG).

RAG es un patrón que acopla un módulo de recuperación, encargado de localizar pasajes relevantes en un corpus, con un modelo generativo que produce la respuesta condicionada por esa evidencia, reduciendo alucinaciones y habilitando trazabilidad mediante citas; su flujo típico abarca ETL/curaduría, indexación, recuperación top-k, re-ranking, prompting con reglas de fidelidad y generación controlada [7]. Por tanto, RAG permite responder preguntas complejas del dominio de compras públicas anclando cada afirmación a fragmentos verificables de pliegos, resoluciones o contratos, requisito central para automatizar pliegos normativos en contextos regulados

e) Recuperación híbrida: búsqueda léxica y vectorial.

La búsqueda léxica prioriza coincidencia exacta y es especialmente eficaz en textos normativos; la recuperación vectorial utiliza embeddings para capturar similitud semántica y cubrir sinonimia/paráfrasis; en la práctica, combinar ambas señales, hibridación, eleva el recall sin degradar en exceso la precisión [8, 9]. La recuperación híbrida es útil en el contexto de compras públicas porque preguntas del tipo “¿qué artículo respalda X?” requieren tanto terminología exacta (léxica) como variantes redaccionales frecuentes en documentos administrativos (vectorial), minimizando el riesgo de omitir evidencia normativa relevante en la elaboración de pliegos.

f) Re-ranking y selección de contexto.

Tras un proceso de recuperación inicial, los re-rankers, específicamente los basados en modelos BERT-like (cross-encoders), son herramientas que se encargan de reordenar los pasajes considerando la interacción entre consulta y texto. Al hacer esto, mejoran la precisión; enfoques de interacción tardía como ColBERT optimizan eficacia con costos controlados [10, 11]; En definitiva, la selección de contexto que verá el generador debe equilibrar diversidad, densidad informativa y límites de ventana para asegurar una respuesta más aterrizada y estructurada, minimizando alucinaciones. En el caso de pliegos electrónicos, re-ranear evidencia normativa asegura priorizar artículos y resoluciones pertinentes, garantizando que el generador utilice el soporte legal más relevante.

g) Prompting, guardrails de fidelidad y citación de fuentes.



El prompting en RAG actúa como mecanismo de control que permite responder únicamente con la evidencia proporcionada, citar (fuente, sección y fecha) y abstenerse ante soporte insuficiente; estas prácticas, presentes en formulaciones canónicas y modelos de lectura-con-fusión (FiD), disminuyen alucinaciones y facilitan auditoría [12, 7]. De esta forma, las plantillas de prompting incorporan políticas de “no respuesta” cuando la evidencia es insuficiente. Además, establecen un formato de citas consistente, indispensables para garantizar que los pliegos generados estén respaldados por citas normativas claras y verificables

h) Métricas y protocolos de evaluación para RAG.

La evaluación en modelos RAG tiene dos componentes, métricas para generación y métricas de recuperación. Para recuperación se emplean métricas como Recall@k, MRR y nDCG los cuales cuantifican, respectivamente, cobertura, rapidez de hallazgo y calidad del ordenamiento de la evidencia recuperada; para generación, se mide fidelidad/groundedness, relevancia y utilidad, manuales y toolkits reconocidos sistematizan estos criterios y marcos automáticos recientes, QAFactEval y RAGAS, instrumentan la verificación factual y la evaluación sin referencia [13, 14]. Para el presente proyecto, estas métricas permitan demostrar, con evidencia cuantitativa, mejoras en cobertura, esto es no perder artículos y/o cláusulas relevantes, fidelidad, cada afirmación es citada, y eficiencia, tiempos de respuesta, frente a flujos manuales de elaboración de pliegos en compras públicas.

5. Metodología de Trabajo

El proyecto se desarrollará con un enfoque iterativo-incremental que permitirá incorporar gradualmente los componentes de recuperación, generación y validación hasta construir un sistema funcional para la generación automática de pliegos en Subasta Inversa Electrónica. En la primera etapa se realizará una revisión exhaustiva del estado del arte en torno a los modelos RAG aplicados a la generación de documentos legales y administrativos, identificando buenas prácticas y limitaciones reportadas en la literatura. De igual manera, se consolidará el marco legal ecuatoriano vigente en materia de Subasta Inversa Electrónica, el cual servirá de referencia para garantizar que los pliegos generados cumplan con la normativa aplicable.

En la segunda etapa, se llevará a cabo la construcción del corpus documental mediante la recopilación y procesamiento de archivos .ushay correspondientes a Subasta Inversa Electrónica en Ecuador. Para el procesamiento de dichos archivos se emplearán técnicas de parsing jerárquico, con herramientas como Docling, que permiten abstraer la estructura XML de manera consistente y generar representaciones utilizables en el pipeline de recuperación. El resultado será un corpus normalizado y listo para indexación.

Luego, con este corpus se diseñará e implementará un pipeline RAG que integre mecanismos de recuperación híbrida, combinando búsqueda léxica y vectorial, estrategias de chunking, embeddings entrenados para el dominio jurídico-administrativo en español y técnicas de re-ranking que prioricen evidencia normativa relevante. Sobre esta base se desarrollará un módulo generativo capaz de redactar pliegos a partir de un prompt con los detalles de la contratación, incorporando guardrails de fidelidad que aseguren el uso exclusivo de evidencia normativa y la citación explícita de las disposiciones legales empleadas.



Una vez implementada la generación automática, se añadirá un módulo de validación dual. En primer lugar, se aplicarán validaciones automáticas enfocadas en consistencia, completitud y cumplimiento normativo. En segundo lugar, se implementará una interfaz gráfica de usuario (GUI) que permitirá a expertos revisar y ajustar manualmente los pliegos generados, visualizando las referencias normativas que respaldan cada sección.

Finalmente, se llevará a cabo una evaluación experimental del sistema mediante métricas de recuperación (Recall, nDCG, precisión de contexto) y métricas de generación (fidelidad normativa, relevancia y utilidad práctica) usando OPIK, una herramienta que permite medir la calidad de las salidas generadas en tareas de RAG, junto con comparaciones frente al flujo manual. Los resultados obtenidos permitirán realizar un análisis de errores y proponer mejoras. Todo el proceso quedará documentado en un informe técnico que incluirá arquitectura, corpus, parámetros de diseño, guía de despliegue y manual de uso de la interfaz gráfica, asegurando la reproducibilidad y continuidad del proyecto.

6. Sumario de Contenidos

1. Introducción

- a. Contexto y problemática en la generación de pliegos de Subasta Inversa Electrónica en Ecuador
- b. Motivación y relevancia del proyecto
- c. Objetivo general y objetivos específicos
- d. Alcance, supuestos y restricciones
- e. Preguntas tipo del dominio y casos de uso
- f. Metodología resumida y contribuciones esperadas

2. Estado del arte

- a. Fundamentos de Retrieval-Augmented Generation (RAG)
- b. Procesamiento de documentos estructurados y jerárquicos (archivos .ushay y XML)
- c. Recuperación híbrida: búsqueda léxica y vectorial
- d. Re-ranking y selección de contexto
- e. Prompting, guardrails de fidelidad y citación de fuentes
- f. Métricas y protocolos de evaluación para RAG (recuperación y generación)
- g. Vacíos identificados y oportunidades para este proyecto

3. Descripción de la Propuesta

- a. Requerimientos funcionales y no funcionales (latencia, trazabilidad, usabilidad)
- b. Arquitectura general del sistema (diagrama)
- c. Pipeline de datos (ETL): ingesta y normalización
- d. Estrategia de recuperación y generación de pliegos
- e. Observabilidad y seguridad: logging y telemetría básica
- f. Plan de evaluación y criterios de aceptación

4. Desarrollo del prototipo

- a. Stack tecnológico y justificación
- b. Construcción del corpus mínimo viable y control de versiones
- c. Implementación del pipeline RAG end-to-end
- d. Parámetros clave (chunking, top-k, re-ranking, temperatura, salidas estructuradas con Pydantic)



- e. Interfaz gráfica de usuario (GUI): diseño, validación manual y trazabilidad normativa
 - f. Pruebas, validación técnica y hardening básico
 - g. Guía de despliegue y operación
 - 5. Experimentos y análisis de resultados
 - a. Diseño experimental y construcción de ground truth
 - b. Métricas de recuperación (Recall@k, MRR/nDCG, Context Precision/Recall)
 - c. Métricas de generación (fidelidad/groundedness, relevancia, utilidad)
 - d. Evaluación con OPIK y comparación con flujos manuales
 - e. Resultados principales
 - f. Estudios de ablación (chunking, top-k, re-ranking, prompting)
 - g. Análisis de errores y casos de estudio
 - h. Discusión: implicaciones, limitaciones y validez de resultados
 - 6. Conclusiones y trabajo futuro
 - a. Síntesis de aportes y aprendizajes
 - b. Limitaciones técnicas y de datos
 - c. Recomendaciones operativas
 - d. Líneas de trabajo futuro
 - 7. Referencias
7. Recursos
- a. Humanos
 - i. Autor: dirección técnica y de proyecto; construcción del corpus; implementación del pipeline RAG (ETL, indexación, recuperación, re-ranking, prompting y guardrails); evaluación y documentación.
Valor: 200 horas \times \$30/h = \$6,000
 - ii. Tutor académico: seguimiento metodológico, validez científica, revisión del documento y del diseño experimental.
Valor: 50 horas \times \$150/h = \$7,500
 - iii. Apoyo técnico puntual: apoyo en creación y curaduría del corpus (normativa, pliegos, contratos, enmiendas), definición de preguntas tipo y validación de respuestas.
Valor: 80 horas \times \$20/h = \$1,600

Subtotal recursos humanos: \$15,100
 - b. Materiales
 - i. Hardware



1. Laptop: MacBook Pro M1 Pro, 16 GB RAM, 512 GB SSD
Valor: \$1,300

2. Google Colab Pro+ (suscripción anual GPU):
Valor: \$1,440

Subtotal hardware: \$2,740

ii. Software

1. Python 3.10+ y librerías open source: \$0
2. Orquestación RAG (LangChain o LlamaIndex): \$0
3. GitHub Pro (control de versiones): \$48
4. Docker Desktop Pro: \$60
5. Milvus (o similares): \$0
6. OPIK (evaluación de RAG): \$0

Subtotal software: \$108

iii. Infraestructura de servidor

1. Servidores en la nube (GPU/embeddings, 200 horas \times \$1/h): \$200

Subtotal servidores: \$200

- i. Formación académica
 - i. Clase Proyecto Integrador (5 créditos): \$1,500

Costo total estimado del proyecto: \$19,648 USD



8. Cronograma de Actividades

Actividades	2025											
	08			09			10			11		
A1: Desarrollo del documento de planificación.												
A2: Revisión del estado del arte y marco normativo aplicable a Subasta Inversa Electrónica.												
A3: Curaduría del corpus												
A4: Procesamiento de archivos .ushay y construcción del corpus mínimo viable.												
A5: Diseño de la arquitectura RAG												
A6: Implementación inicial del pipeline RAG (indexación léxica y vectorial, re-ranking, prompting con guardrails)												
A7: Desarrollo de la capa generativa de pliegos y salidas estructuradas.												
A8: Implementación de validaciones automáticas (consistencia, completitud, referencias normativas).												
A9: Desarrollo de la interfaz gráfica de usuario (GUI) para validación manual.												
A10: Evaluación experimental con métricas de recuperación y generación.												
A11: Integración de OPIK para evaluación automatizada y comparación con flujo manual.												
A12: Análisis de errores y estudios de ablación.												





Entregable 3 – 23/11/2025

Contenido:

- Validaciones automáticas implementadas (consistencia, completitud y referencias normativas).
- Interfaz gráfica de usuario (GUI) para validación manual (versión funcional).
- Evaluación experimental inicial con métricas de recuperación y generación.
- Reporte comparativo de evaluación con OPIK y flujo manual.

Formato de entrega: Documento PDF con resultados, prototipo funcional (pipeline + GUI) en repositorio GitHub.

Entregable Final – 14/12/2025

Contenido:

- Análisis de errores y estudios de ablación.
- Documentación técnica completa (arquitectura, diseño, guía de despliegue, manual de uso de la interfaz).
- Documento de conclusiones, limitaciones y recomendaciones.
- Repositorio final con: código fuente completo del pipeline RAG y la interfaz gráfica, corpus procesado (.ushay + normativa), scripts de evaluación (incluyendo OPIK), documentación técnica y científica.

Formato de entrega: Documento PDF final, repositorio GitHub completo, presentación en PowerPoint/PDF.



10. Referencias

- [1] E. K. a. M. Fazekas, «Hidden barriers to open competition: Using text mining to uncover corrupt restrictions in public procurement,» 2024. [En línea]. Available: https://www.govtransparency.eu/wp-content/uploads/2024/05/Katona-Fazekas_HU_text-miningPP-corruption240417_GTIpublish240430final.pdf .
- [2] ECPR, «Unveiling Corrupt Practices in Public Procurement Across Europe: A Text Mining Approach,» 2023. [En línea]. Available: <https://ecpr.eu/Events/Event/PaperDetails/75709>.
- [3] E. D. P. P. N. S. Cerqua, «Corruption red flags in public procurement: new evidence from Italian roadworks contracts,» *EPJ Data Science*, 2022.
- [4] H. K. School, «Overcoming Corruption and War—Lessons from Ukraine’s ProZorro Procurement System,» *Harvard Kennedy School*, 2023.
- [5] «DOZORRO: Artificial Intelligence to Find Violations in ProZorro—How It Works,» *Transparency International Ukraine*, 2018.
- [6] N. Pipitone y G. H. Alami, «LegalBench-RAG: A Benchmark for Retrieval-Augmented Generation in the Legal Domain,» 2024. [En línea]. Available: <https://arxiv.org/pdf/2408.10343>.
- [7] E. P. P. Lewis, F. P. A. Piktus, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel y D. Kiela, «Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,» *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 2020. [En línea]. Available: proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.
- [8] S. Robertson y H. Zaragoza, «The Probabilistic Relevance Framework: BM25 and Beyond,» *Foundations and Trends in Information Retrieval*,» *staff.city.ac.uk* , vol. 3, n° 4, pp. 333-389, 2009.
- [9] V. Karpukhin, B. Oğuz, S. Min y P. Lewis, «Dense Passage Retrieval for Open-Domain Question Answering,» *Proc. EMNLP*, 2020. [En línea]. Available: aclanthology.org/2020.emnlp-main.550.pdf.
- [10] R. Nogueira y K. Cho, «Passage Re-Ranking with BERT,» *arXiv*, 2019. [En línea]. Available: arxiv.org/abs/1901.04085.
- [11] O. Khattab y M. Zaharia, «ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT,» *arXiv*, 2020. [En línea]. Available: arxiv.org/abs/2004.12832.
- [12] G. Izacard y É. Grave, «Leveraging Passage Retrieval with Generative Models for Open-Domain QA (FiD),» *ACL Anthology*, 2021. [En línea]. Available: aclanthology.org/2021.eacl-main.74.pdf.
- [13] C. D. Manning, P. Raghavan y H. Schütze, «Introduction to Information Retrieval. Cambridge: Cambridge,» Univ. Press, [En línea]. Available: nlp.stanford.edu/IR-book.



- [14] E. Shahul, L. Espinosa-Anke y J. James, «RAGAS: Automated Evaluation of Retrieval Augmented Generation,» Demos, 2024. [En línea]. Available: aclanthology.org/2024.eacl-demo.16.pdf.

11. Revisión y firma del tutor del proyecto

Yo, Daniel Riofrío, profesor de la carrera de Ingeniería en Ciencias de la Computación, hago constar que he revisado y, por lo tanto, apruebo el documento de planificación del proyecto titulado “Modelo de Recuperación Aumentada por Generación (RAG) para la Generación Automática de Pliegos en Compras Públicas de Subasta Inversa Electrónica en Ecuador.” propuesto por el estudiante Pablo Andrés Encalada Andrade. Por otra parte, me comprometo a proporcionar al estudiante el soporte necesario y oportuno para el buen desarrollo del proyecto antes mencionado.



Fdo: Daniel Riofrío

Quito, 5 de septiembre de 2025