



**Agente de inteligencia artificial para la elaboración asistida de respuestas a peticiones
ciudadanas en procesos de selección por mérito de la CNSC**

Juan Sebastian Ballen Villalba

Universidad Ean
Facultad de ingeniería
Maestría en Ciencia de Datos
Bogotá, Colombia
25 de junio del 2026

**Agente de inteligencia artificial para la elaboración asistida de respuestas a peticiones
ciudadanas en procesos de selección por mérito de la CNSC**

Juan Sebastian Ballen Villalba

Magister en Ciencia de Datos

Director:

Leonardo Pérez Cortes

Modalidad:

Trabajo Dirigido

Juan Sebastian Ballen Villalba

Universidad Ean

Facultad de ingeniería

Maestría en Ciencia de Datos

Bogotá, Colombia

25 de junio del 2026

Nota de aceptación:

Firma del jurado

Firma del jurado

Firma del director del trabajo de grado

Bogotá, 25 de junio del 2026

Resumen

El presente trabajo de grado aborda el diseño y desarrollo de un agente de inteligencia artificial orientado a apoyar la elaboración de respuestas a peticiones ciudadanas en procesos de selección por mérito administrados por la Comisión Nacional del Servicio Civil (CNSC). El agente se fundamenta en la arquitectura de Generación Aumentada por Recuperación (RAG), que permite consultar un repositorio normativo versionado mediante lenguaje natural, recuperar fragmentos pertinentes con citación estructurada y generar borradores de respuesta con citas trazables al documento fuente. La investigación es de tipo aplicada, con enfoque mixto y alcance exploratorio-descriptivo. Se empleó la metodología DOFA para el análisis externo y un modelo diagnóstico de madurez estructurado en siete dimensiones para el análisis interno. El prototipo funcional fue desplegado como servicio web y validado mediante un piloto controlado ejecutado entre el 13 y el 30 de abril de 2026 con funcionarios de la entidad. La validación del desempeño se realizó mediante una encuesta estructurada tipo Likert de 25 ítems ($N = 5$) y un cuestionario de preguntas abiertas ($n = 2$). Los resultados evidenciaron una valoración favorable de la pertinencia y vigencia de las citas normativas y de la utilidad del borrador como punto de partida, junto con una valoración baja del desempeño temporal del prototipo, atribuible a la ejecución del modelo de lenguaje exclusivamente sobre CPU. El proyecto contribuye a la estandarización de la argumentación jurídica, la trazabilidad de las citas normativas y la reducción del esfuerzo en la elaboración de respuestas, manteniendo la revisión humana obligatoria como condición previa al envío de cualquier comunicación oficial.

Palabras clave: inteligencia artificial, RAG, peticiones ciudadanas, servicio civil, procesamiento de lenguaje natural, CNSC, recuperación de información.

Abstract

This thesis addresses the design and development of an artificial intelligence agent aimed at supporting the drafting of responses to citizen petitions in merit-based selection processes managed by the National Civil Service Commission (CNSC) of Colombia. The agent is based on the Retrieval-Augmented Generation (RAG) architecture, which enables querying a versioned regulatory repository using natural language, retrieving relevant fragments with structured citations, and generating draft responses with traceable references to the source document. The research follows an applied methodology with a mixed approach and an exploratory-descriptive scope. A SWOT methodology was used for external analysis, alongside a seven-dimension maturity diagnostic model for internal analysis. The functional prototype was deployed as a web service and validated through a controlled pilot conducted between April 13 and April 30, 2026, with the entity staff. Performance validation was carried out through a structured 25-item Likert-type survey (N = 5) and an open-ended questionnaire (n = 2). The results showed a favorable assessment of the relevance and validity of the normative citations and of the usefulness of the draft as a starting point, together with a low assessment of the prototype's response time, attributable to running the language model on CPU only. The project contributes to the standardization of legal argumentation, traceability of regulatory citations, and reduction of effort in response drafting, while maintaining mandatory human review prior to sending any official communication.

Keywords: artificial intelligence, RAG, citizen petitions, civil service, natural language processing, CNSC, information retrieval.

Contenido

Introducción	12
Objetivos	13
Objetivo general.....	13
Objetivos específicos	13
Justificación	14
Marco Institucional.....	16
Presentación general de la entidad.....	16
Referentes estratégicos	18
Estructura organizacional	18
Productos y servicios ofertados	21
Análisis del sector	22
Marco de Referencia	22
Normatividad que usa el agente	23
Recuperación de información	23
Procesamiento de Lenguaje Natural	24
Generación Aumentada por Recuperación (RAG)	24
Fundamentación matemática de la búsqueda híbrida y de las arquitecturas RAG	25
Búsqueda densa, búsqueda dispersa y su complementariedad.....	25
Arquitectura Transformer y modelos de lenguaje pequeños (SLM)	26
Métricas de evaluación de arquitecturas RAG	27
Comparación de enfoques.....	27
Gobernanza de datos y seguridad de la información	28
Componente sociotécnico.....	28
Antecedentes institucionales y técnico-organizacionales	28
Brechas identificadas.....	29
Articulación entre referentes teóricos y la solución propuesta.....	29

Diseño Metodológico	31
Tipo de investigación	31
Análisis externo	31
Análisis interno	32
Población, muestra y ficha técnica.....	33
Identificación de variables.....	34
Instrumento de medición.....	35
Validación del instrumento.....	35
Diagnóstico Organizacional	36
Línea base operativa	36
Situación actual.....	40
Fortalezas identificadas	40
Oportunidades de mejora.....	40
Procesamiento estadístico de datos.....	41
Análisis de los resultados.....	41
Perfil de los participantes y confiabilidad del instrumento	41
Resultados cuantitativos de la encuesta Likert.....	43
Perfil global y por dimensión	43
Resultados a nivel de ítem: aspectos mejor y peor valorados	45
Variabilidad entre participantes	47
Análisis cualitativo del cuestionario de preguntas abiertas.....	47
Experiencia de uso del agente durante el piloto	48
Manejo de peticiones multi-pregunta.....	48
Impacto en la eficiencia y la calidad del proceso.....	48
Barreras, limitaciones y oportunidades de mejora	48
Valoración general y recomendaciones	49
Triangulación y discusión de hallazgos	49

Plan de Intervención	50
Arquitectura de la solución.....	50
Stack tecnológico.....	52
Justificación de la selección del modelo de lenguaje y la arquitectura de inferencia	53
Protocolo del piloto	57
Entregables del plan de intervención	57
Plan de implementación (propuesta para producción)	57
Fases y cronograma	58
Roles y responsabilidades	58
Requisitos y gestión del cambio.....	58
Punto de partida: lo ejecutado en el piloto.....	59
Conclusiones y Recomendaciones	59
Conclusiones	59
Recomendaciones	61
Modelo de operación, transferencia y sostenibilidad	62
Modelo de operación post-piloto	62
Escalabilidad y replicabilidad hacia otras entidades	62
Propiedad intelectual y licenciamiento	63
Modelo de costos operativos anualizados.....	63
Referencias	64
Lista de Anexos	68
Anexo A - Carta de autorización para la intervención empresarial	69
Anexo B - Datos de peticiones por etapa del proceso de selección DIAN 2022	70
Anexo C - Reporte de control de términos (período 250820 al 251020).....	71
Anexo D - Cuestionario de percepción del Agente de IA.....	72
Anexo E - Encuesta de evaluación del Agente de IA (tipo Likert).....	74

Anexo F - Repositorio del código fuente del agente	77
--	----

Tabla de figuras

Figura 1 Filosofía institucional de la CNSC. Elaboración propia.....	18
Figura 2 Estructura organizacional. Elaboración propia a partir del organigrama de la CNSC	20
Figura 3 Sub-árbol operativo del proceso de atención de peticiones. Elaboración propia a partir de la estructura oficial de la CNSC (Acuerdo 2073 de 2021).....	21
Figura 4 Matriz DOFA, elaboración propia	32
Figura 5 Mapa del proceso actual para la elaboración de respuestas a peticiones en la CNSC. Elaboración propia.	38
Figura 6 Proceso desde la creación de la petición hasta la respuesta al peticionario, con tiempos promedio por hito. Elaboración propia a partir de datos tomados de la CNSC.....	39
Figura 7 Distribución global de respuestas en la escala Likert. Elaboración propia a partir de los datos del piloto	43
Figura 8 Perfil de medias por dimensión (diagrama radar, escala 1–5). Elaboración propia a partir de los datos del piloto.....	44
Figura 9 Medias por dimensión del instrumento de evaluación. Elaboración propia a partir de los datos del piloto.	45
Figura 10 Medias por ítem de la encuesta de evaluación. Elaboración propia a partir de los datos del piloto.	46
Figura 11 Arquitectura del agente RAG para la elaboración asistida de respuestas a peticiones. Elaboración propia.	56

Tabla de tablas

Tabla 1 Vínculos entre conceptos teóricos y componentes de la solución propuesta. Elaboración propia a partir de los referentes teóricos consultados.....	31
Tabla 2 Ficha técnica de la investigación. Elaboración propia	34
Tabla 3 Tiempos promedio por hito del proceso de atención de peticiones (periodo agosto-octubre 2025). Elaboración propia a partir de datos tomados de la CNSC (2025c).	37
Tabla 4 Coeficiente Alfa de Cronbach por dimensión y global. Elaboración propia a partir de los datos del piloto.	42
Tabla 5 Medias y desviaciones estándar por dimensión. Elaboración propia a partir de los datos del piloto.	45
Tabla 6 Stack tecnológico del agente. Elaboración propia a partir de la configuración del prototipo funcional.....	53
Tabla 7 Comparación de alternativas de modelo e inferencia según latencia, costo y privacidad. Elaboración propia.	55
Tabla 8 Roles y responsabilidades para la operación del agente. Elaboración propia...58	
Tabla 9 Modelo de costos operativos anualizados (referencial). Elaboración propia a partir de Oracle (2026).	63

Introducción

La gestión del empleo público por mérito en Colombia exige a las entidades responsables actuaciones estandarizadas, fundamentadas en normatividad vigente y plenamente trazables. La Comisión Nacional del Servicio Civil (CNSC), como órgano autónomo encargado de administrar y vigilar la carrera administrativa, gestiona convocatorias de selección que generan un volumen significativo de peticiones ciudadanas cuya respuesta debe cumplir con los principios de oportunidad, completitud y respuesta de fondo, consagrados en la Ley 1755 de 2015 (Congreso de la República, 2015). En este contexto, el presente trabajo se sitúa en la intersección entre la gestión pública del servicio civil, la ciencia de datos y la inteligencia artificial aplicada al procesamiento de información jurídico-documental, dentro de la línea de investigación en Tecnología de la Información y Comunicaciones de la Universidad Ean.

El proceso actual de respuesta a peticiones en la CNSC presenta cuellos de botella que comprometen la calidad y oportunidad del servicio. Una vez la petición es recibida, verificada y asignada, el funcionario responsable debe identificar las solicitudes del ciudadano, contrastar información del aspirante en fuentes como el Sistema de Información y Manejo de Empleo Público (SIMO) o reportes estructurados, y localizar el soporte normativo vigente aplicable entre múltiples fuentes, tales como leyes, decretos, acuerdos de convocatoria, anexos y lineamientos, para redactar una respuesta con fundamentación jurídica precisa. Este esfuerzo depende en gran medida de la experiencia individual del funcionario y genera riesgos de inconsistencia en la argumentación, variabilidad en las citas normativas y retrabajos por devoluciones en las etapas de revisión jurídica o firma. Las métricas internas del periodo comprendido entre el 20 de agosto y el 20 de octubre de 2025, correspondientes al Grupo de Gerencia de Convocatoria 12 del Proceso de Selección DIAN 2676, evidencian que el tramo más exigente del ciclo corresponde a la elaboración de la respuesta, con un promedio de 3,4 días hábiles, el mayor de todos los hitos, seguido de la revisión jurídica (2,5 días), la firma (1,0 día) y la asignación (0,4 días), sobre un tiempo total promedio de 6,5 días hábiles desde la radicación hasta la comunicación al peticionario (CNSC, 2025c). A título ilustrativo, el Proceso de Selección DIAN 2022, proceso anterior administrado por la entidad, atendió 20.162 peticiones ciudadanas, cifra que permite dimensionar la presión operativa estructural sobre las dependencias responsables, con independencia del proceso específico considerado.

Ante esta problemática, se plantea la siguiente pregunta de investigación: ¿en qué medida un agente interno de inteligencia artificial, orientado a la recuperación verificable de

normatividad vigente, puede apoyar la elaboración de borradores de respuesta y cómo perciben los funcionarios su utilidad, su efecto sobre el tiempo de elaboración y la calidad de las citas normativas, en comparación con el proceso manual actual en la CNSC? Esta formulación es coherente con el alcance exploratorio-descriptivo del estudio, que caracteriza la percepción y el desempeño del prototipo sin pretender establecer una relación causal mediante un diseño experimental antes-después. Para responder esta pregunta, el proyecto propone el desarrollo de un agente de IA jurídica basado en la arquitectura de Generación Aumentada por Recuperación (RAG), que integre consulta en lenguaje natural sobre un repositorio normativo versionado, recuperación de fragmentos con citación estructurada y generación de borradores de respuesta con citas trazables al documento fuente, manteniendo en todo momento la revisión humana obligatoria como condición previa al envío de cualquier comunicación oficial (Lewis et al., 2020; Klesel y Wittmann, 2025).

El documento se organiza en las siguientes secciones. En primer lugar, se presentan los objetivos, tanto el general como los específicos. Posteriormente se desarrolla la justificación del proyecto, seguida del marco institucional donde se describe la CNSC como escenario de intervención. El marco de referencia fundamenta teóricamente la propuesta desde la recuperación de información, el procesamiento de lenguaje natural, la arquitectura RAG y los principios de IA confiable en gobierno. Luego se detalla el diseño metodológico, que comprende el enfoque de investigación, el modelo diagnóstico, los instrumentos de recolección y el protocolo de validación. Finalmente se presentan las conclusiones y recomendaciones, las referencias bibliográficas y los anexos correspondientes.

Objetivos

Objetivo general

Desarrollar y validar un agente de inteligencia artificial de apoyo a los funcionarios responsables en la elaboración de borradores de respuesta a peticiones ciudadanas de procesos de selección por mérito en la CNSC, que integre la consulta en lenguaje natural de normatividad vigente y el contraste con datos del aspirante, para producir borradores sustentados con citas verificables y trazables al documento fuente, bajo revisión humana obligatoria.

Objetivos específicos

- Consolidar un repositorio normativo controlado y versionado aplicable al derecho de petición y a los procesos de selección por mérito de la CNSC, incorporando metadatos de fuente, fecha, estado de vigencia, numeral e identificador, junto con una guía de actualización y trazabilidad que asegure la consulta reproducible.
- Implementar un componente de Generación Aumentada por Recuperación (RAG) que indexe el repositorio normativo, recupere fragmentos pertinentes con citación estructurada y aplique reglas de control de vigencia sobre cada fragmento recuperado, garantizando que únicamente se cite normatividad vigente y que cada referencia sea trazable al documento fuente.
- Implementar un componente de comprensión de la petición y de generación asistida del borrador que, a partir del texto de la solicitud, segmente las sub-solicitudes del ciudadano (peticiones multi-pregunta) y produzca, sobre la evidencia normativa recuperada por el componente RAG, un borrador estructurado con citas explícitas, manteniendo el enfoque de apoyo al funcionario y la revisión humana obligatoria.
- Estandarizar el formato de salida del agente mediante plantillas institucionales que integren borrador argumentado, citas verificables y evidencias normativas recuperadas, incluyendo campos parametrizables, numeración, estilos institucionales, trazabilidad a la fuente y exportación a formato documental.
- Validar el desempeño del agente mediante un piloto controlado de carácter exploratorio, aplicando una encuesta estructurada tipo Likert y un cuestionario de preguntas abiertas a los funcionarios que elaboran las respuestas, para caracterizar la utilidad percibida, la percepción sobre el tiempo de elaboración del borrador y la calidad de las citas normativas generadas, sobre un conjunto de casos reales del proceso bajo criterios de evaluación predefinidos.

Justificación

El proyecto propone el desarrollo de un asistente interno para la elaboración de borradores de respuesta a peticiones ciudadanas en procesos de selección por mérito, que automatiza tareas como el cruce con la base del aspirante en la plataforma SIMO o en archivos estructurados (CSV), y la redacción con soporte normativo vigente. Su relevancia se fundamenta en el cumplimiento de las exigencias legales relacionadas con la oportunidad, la integridad y la respuesta sustancial del derecho de petición, conforme a lo establecido en la Ley 1755 de 2015 (Congreso de la República, 2015). Asimismo, responde a la necesidad de

asegurar uniformidad, trazabilidad y mérito dentro del proceso de provisión de empleos públicos (Ley 909 de 2004; Decreto 1083 de 2015). En los procesos masivos como los que administra la CNSC para concursos de mérito, estas exigencias generan una presión operativa significativa y un riesgo de variabilidad cuando la gestión permanece enteramente manual.

En términos de beneficios esperados para la entidad, la sustitución de la lectura y búsqueda manual por extracción y recuperación asistida proyecta una reducción del tiempo de elaboración de las respuestas y una mayor consistencia argumentativa, en la medida en que un mismo tipo de caso se resuelva con la misma norma, el mismo numeral y la misma vigencia. Esto eleva la calidad del servicio al ciudadano y la confiabilidad institucional. La evidencia internacional sobre IA en el sector público indica que los mayores retornos tempranos se logran en casos de uso documentales y de servicio al ciudadano, debido a su alto volumen y estandarización, siempre que se implementen con trazabilidad y bajo control humano (Neumann et al., 2022; Straub et al., 2023).

Desde el punto de vista metodológico, la Generación Aumentada por Recuperación (RAG) es un enfoque consolidado para generar textos anclados en evidencia recuperable, minimizando la alucinación y la desactualización al obligar a citar pasajes exactos del corpus versionado antes de redactar (Lewis et al., 2020). Como referencia operativa, la línea base registrada entre el 20 de agosto y el 20 de octubre de 2025 para el Grupo de Gerencia de Convocatoria 12 del Proceso de Selección DIAN 2676 indica un tiempo promedio total de 6,5 días hábiles desde la creación de la petición hasta la contestación al peticionario y un tiempo promedio de elaboración de 3,4 días por caso, entendiendo la elaboración como el tramo desde la asignación al funcionario hasta el envío a revisión jurídica (CNSC, 2025c). Estos datos evidencian que la elaboración del borrador concentra poco más de la mitad del ciclo y constituye el punto donde la intervención tecnológica puede generar mayor impacto, con independencia de la etapa del proceso de selección en la que se aplique. El efecto real de la solución sobre estos tiempos y sobre la percepción de utilidad de los funcionarios se evaluó mediante una encuesta estructurada aplicada durante el piloto, conforme se detalla en el diseño metodológico.

En cuanto a la sostenibilidad técnica, el índice vectorial FAISS permite una búsqueda eficiente y escalable sobre la normativa, con infraestructura de bajo costo y soporte maduro en Python, lo cual facilita la adopción y el mantenimiento por equipos internos de la entidad. La alineación estratégica del proyecto se articula con los principios de legalidad, economía, igualdad y celeridad del régimen de función pública, y con marcos contemporáneos de IA confiable en el sector público, al incorporar controles de vigencia, versionado y revisión

obligatoria de los borradores antes del envío (OCDE, 2019; NIST, 2023). Esto habilita una ruta de mejora continua aplicable a las distintas etapas del proceso de selección, tales como verificación de requisitos mínimos, pruebas y listas de elegibles.

En el contexto del sector público y de la CNSC, la privacidad de los datos de los ciudadanos y la reserva legal de ciertos documentos son condiciones críticas del diseño. Por ello, la elección de una arquitectura de código abierto y de ejecución local, en lugar del envío de información a interfaces de programación (API) comerciales externas como las de OpenAI o Anthropic, responde a la necesidad de garantizar la soberanía del dato, la seguridad de la información y el cumplimiento de la normativa de protección de datos personales (Ley 1581 de 2012). Esta decisión asegura que las peticiones y los datos del aspirante se procesen dentro de la infraestructura controlada por el proyecto, sin transmitir información gubernamental a terceros. A lo anterior se suma que el prototipo se construyó sobre la capa gratuita de un proveedor de nube, dado que la entidad no cuenta con infraestructura tecnológica especializada para este tipo de soluciones; esta restricción, adoptada por viabilidad, limita la capacidad de cómputo disponible y explica en buena parte los tiempos de respuesta percibidos como bajos durante el piloto, aspecto que se retoma en el análisis de resultados y en las recomendaciones.

Desde la perspectiva del ciudadano, el proyecto aporta valor público directo al mejorar la oportunidad, la claridad y la previsibilidad de las respuestas de la administración. La estandarización de las citas normativas y la incorporación de trazabilidad en las fuentes citadas permitirá al peticionario verificar el fundamento jurídico de la decisión, fortaleciendo la transparencia del proceso.

El presente trabajo se inscribe en el campo de investigación de Ciencia, Tecnología e Innovación de la Universidad Ean, dentro de la línea de Tecnología de la Información y Comunicaciones, y se desarrolla en la Comisión Nacional del Servicio Civil como escenario de intervención.

Marco Institucional

Presentación general de la entidad

La Comisión Nacional del Servicio Civil (CNSC) es un órgano autónomo e independiente, del más alto nivel en la estructura del Estado colombiano, dotado de personería jurídica, autonomía administrativa, patrimonial y técnica, que no hace parte de ninguna de las ramas del poder público (Congreso de la República, 2004). Su creación tiene origen en el

artículo 130 de la Constitución Política de 1991, que establece que habrá una Comisión Nacional del Servicio Civil responsable de la administración y vigilancia de las carreras de los servidores públicos. Sin embargo, la entidad solo entró en funcionamiento efectivo con la promulgación de la Ley 909 de 2004, que reglamentó el empleo público, la carrera administrativa y la gerencia pública en Colombia, y definió las competencias de la CNSC como órgano de garantía y protección del sistema de mérito (Congreso de la República, 2004).

Históricamente, la necesidad de un órgano garante del mérito en el acceso al empleo público se remonta a los debates constituyentes de 1991, cuando se reconoció que la administración pública requería un mecanismo independiente para evitar que el clientelismo y las prácticas informales desplazaran los criterios técnicos en la provisión de cargos. Aunque la Constitución creó la figura de la CNSC, transcurrieron más de diez años antes de que la entidad contara con un marco legal operativo, pues fue la Ley 909 de 2004 la que definió su estructura, competencias y procedimientos para la administración de los concursos de mérito. A partir de ese momento, la CNSC asumió la responsabilidad de diseñar, convocar y administrar los procesos de selección para las entidades del orden nacional y territorial que pertenecen al sistema general y a los sistemas específicos de carrera administrativa.

La misión de la CNSC consiste en administrar y vigilar la carrera administrativa de los servidores públicos, garantizando el mérito como principio rector del ingreso y el ascenso en el empleo público colombiano. Su visión se orienta a consolidarse como la entidad de referencia en la profesionalización del servicio público en Colombia, promoviendo la transparencia, la igualdad de oportunidades y la excelencia en la gestión del talento humano del Estado. Estos referentes orientan la totalidad de sus actuaciones administrativas, desde la expedición de acuerdos de convocatoria hasta la atención de las peticiones que los ciudadanos presentan durante las distintas etapas de los procesos de selección (CNSC, s. f.).

Desde su puesta en marcha, la CNSC ha administrado más de 87 procesos de selección a través de los cuales cerca de 311.000 colombianos han accedido a empleos de carrera administrativa o han ascendido en entidades públicas de todo el país (CNSC, 2022). Estas cifras reflejan el papel central que desempeña la entidad en la profesionalización del servicio público y en la materialización del principio constitucional del mérito como criterio de acceso al empleo estatal. A título ilustrativo, solo en el Proceso de Selección DIAN 2022 se atendieron 20.162 peticiones ciudadanas distribuidas entre las etapas de inscripciones, verificación de requisitos mínimos, pruebas y conformación de listas de elegibles, lo cual permite dimensionar la escala operativa de la entidad y la presión que recae sobre los equipos responsables de elaborar las respuestas.

En este proyecto, la CNSC constituye el escenario de intervención donde se desarrolla e implementa el agente de inteligencia artificial para la elaboración asistida de respuestas a peticiones. La entidad gestiona convocatorias de selección, verificación de requisitos mínimos, aplicación de pruebas y conformación de listas de elegibles, actividades que generan un volumen significativo de comunicaciones y peticiones ciudadanas que deben resolverse con oportunidad, completitud y respuesta de fondo, conforme a la Ley 1755 de 2015 (Congreso de la República, 2015). El proyecto busca optimizar la elaboración de estas respuestas mediante la extracción automatizada de la información contenida en las peticiones, el cruce con la base de datos del aspirante y la redacción asistida con citas normativas vigentes y verificables.

Referentes estratégicos

La CNSC se rige por los principios de legalidad, transparencia, igualdad y mérito consagrados en la Ley 909 de 2004 y desarrollados en el Decreto 1083 de 2015, que exigen actuaciones homogéneas, trazables y sujetas a reglas (Congreso de la República, 2004; Presidencia de la República, 2015). En coherencia con el Modelo Integrado de Planeación y Gestión (MIPG) y la Política de Gobierno Digital, se promueven la estandarización de procesos, la simplificación de trámites, la gestión documental y la trazabilidad de decisiones. Bajo estos referentes, el proyecto se alinea con las metas institucionales de reducir tiempos de ciclo, mejorar la consistencia argumentativa y fortalecer la trazabilidad de las respuestas, en armonía con el derecho fundamental de petición (Congreso de la República, 2015) y con las políticas de mejora de la gestión pública.



Figura 1 Filosofía institucional de la CNSC. Elaboración propia

Estructura organizacional

La CNSC cuenta con una estructura formal definida en el Acuerdo 2073 de 2021, que organiza la entidad en Sala Plena de Comisionados, Presidencia, Despachos de Comisionados, Secretaría General y dependencias técnicas y de apoyo, entre otras (CNSC, 2021). La Sala Plena, compuesta por tres comisionados elegidos por concurso de méritos para periodos institucionales de cuatro años, constituye la máxima instancia de decisión y es responsable de expedir los acuerdos de convocatoria que rigen cada proceso de selección.

Para efectos del proyecto, los actores operativos directamente involucrados son: Atención al Ciudadano (recepción y radicación de peticiones), Gestión Documental (clasificación, archivo y trazabilidad), Tecnología SIMO (plataforma que contiene la información del aspirante inscrito en cada convocatoria), el área Jurídica (emisión de lineamientos normativos y revisión de fondo de los borradores de respuesta) y el Equipo del Proyecto (operación del prototipo, recolección de indicadores y acciones de mejora). Este sub-árbol operativo soporta el flujo del piloto y habilita responsabilidades claras y trazabilidad en cada etapa del proceso.

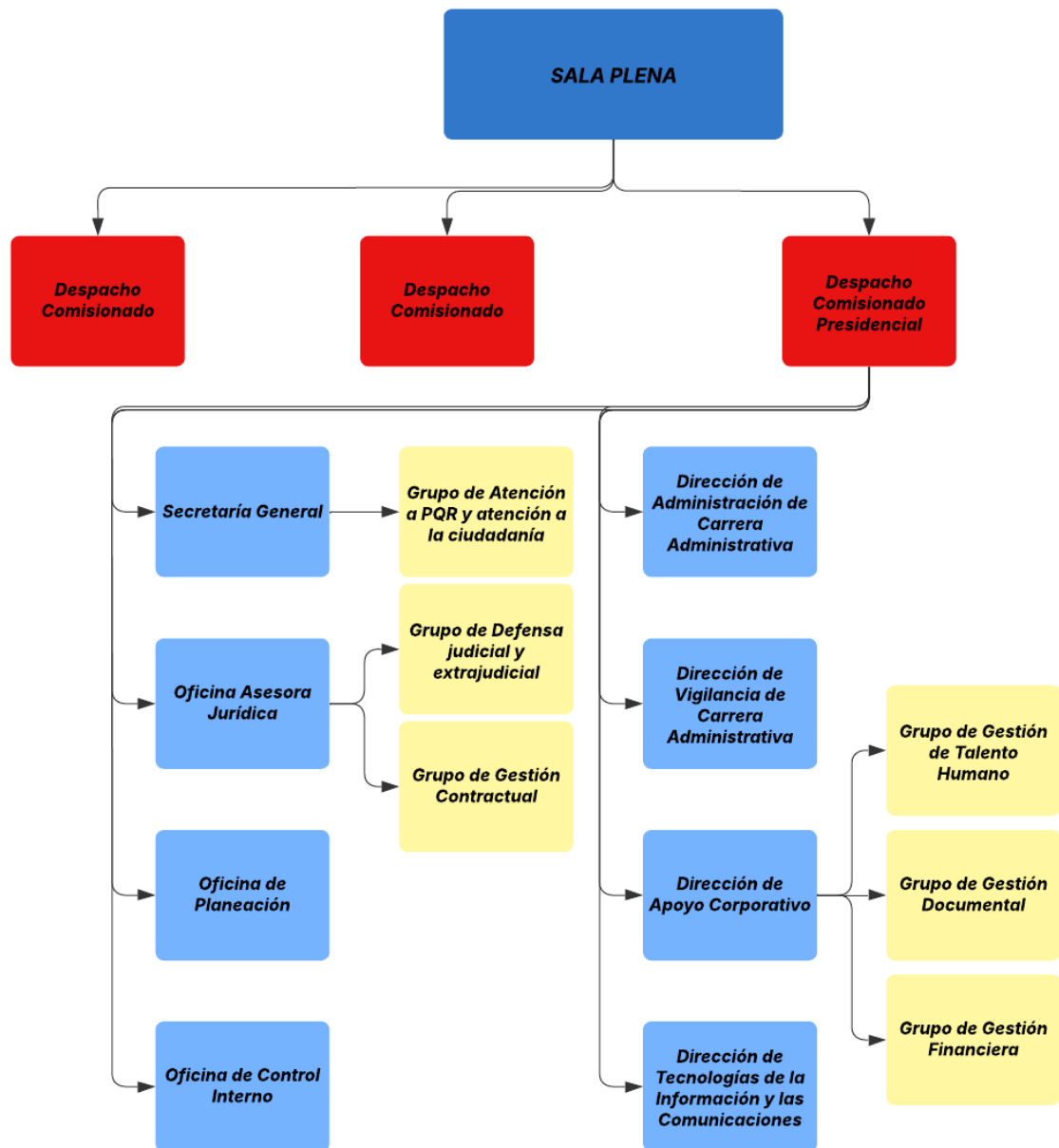


Figura 2 Estructura organizacional. Elaboración propia a partir del organigrama de la CNSC

La ilustración muestra el sub-árbol operativo del proceso de atención de peticiones, donde se detallan los actores y su rol funcional dentro del flujo.



Figura 3 Sub-árbol operativo del proceso de atención de peticiones. Elaboración propia a partir de la estructura oficial de la CNSC (Acuerdo 2073 de 2021).

Este sub-árbol operativo delimita los puntos de integración del agente (extracción de la solicitud del aspirante, cruce con SIMO/CSV y recuperación normativa con control de vigencias), los controles de calidad y trazabilidad (metadatos, bitácoras y revisión jurídica), sin alterar la jerarquía institucional.

Productos y servicios ofertados

La CNSC ofrece procesos de selección por mérito para entidades del Estado, que incluyen la divulgación de convocatorias, las inscripciones, la verificación de requisitos mínimos, la aplicación de pruebas, la consolidación y publicación de resultados y la conformación de listas de elegibles, además de la atención de peticiones, quejas y reclamos asociadas a cada etapa. Cada convocatoria se rige por un acuerdo y un anexo técnico que establecen las reglas del proceso, los requisitos por empleo, los cronogramas y los criterios de evaluación. La complejidad normativa es considerable, dado que un solo proceso de selección

puede involucrar decenas de empleos con requisitos diferenciados, lo que multiplica las fuentes que el funcionario debe consultar al momento de elaborar cada respuesta.

El asistente propuesto agrega valor a estos servicios al estandarizar la extracción de la parte petitoria, vincular datos del aspirante desde SIMO o archivos estructurados (CSV) y fundamentar los borradores de respuesta en normativa vigente con citas verificables y control de vigencia, manteniendo revisión humana previa al envío.

Análisis del sector

El proyecto se ubica en el sector público y, en particular, en la gestión del talento humano en el Estado. El marco regulatorio vigente (Ley 909 de 2004; Decreto 1083 de 2015) demanda actuaciones uniformes y auditables, mientras que la Ley 1755 de 2015 obliga a responder peticiones en términos de oportunidad y de fondo. Las tendencias de digitalización gubernamental y mejora regulatoria impulsan soluciones de procesamiento documental y analítica jurídica que reduzcan variabilidad y tiempos de gestión, tendencia que se ha acelerado en los últimos años con la Política de Gobierno Digital (MinTIC, 2022) y con las recomendaciones de la OCDE sobre uso responsable de inteligencia artificial en el sector público (OCDE, 2019).

En el contexto colombiano, varias entidades del Estado han avanzado en la implementación de herramientas de automatización documental y gestión del conocimiento normativo. Sin embargo, la integración de tecnologías de recuperación semántica con generación de texto asistida por IA es aún incipiente en el ámbito de la gestión de peticiones ciudadanas, lo que posiciona a este proyecto como una iniciativa de frontera en el sector. La disponibilidad de portales de Transparencia y repositorios normativos oficiales favorece la construcción de un corpus versionado y la adopción de arquitecturas RAG con búsqueda vectorial (FAISS), que anclan cada respuesta a pasajes normativos trazables. En este entorno, un asistente jurídico-documental contribuye a la eficiencia operativa y a la confianza ciudadana, con escalabilidad hacia otras etapas del proceso de selección.

Marco de Referencia

El empleo público por mérito en Colombia se fundamenta en el principio de igualdad de oportunidades y en la exigencia de actuaciones homogéneas y trazables, lo que impone a la CNSC estándares de calidad en la gestión de información y en la respuesta a peticiones asociadas a cada proceso de selección. Este marco normativo está dado, entre otros, por la

Ley 909 de 2004 y el Decreto 1083 de 2015, así como por la Ley 1755 de 2015 sobre derecho de petición, que demanda respuestas claras, dentro de los términos legales y fundadas, con pleno respeto por el debido proceso y la motivación administrativa (Congreso de la República, 2004, 2015; Presidencia de la República, 2015).

En este contexto, hay etapas del proceso de selección que concentran un alto volumen de consultas y solicitudes, las cuales requieren estandarización, trazabilidad y gestión documental robusta para sostener decisiones a lo largo del ciclo. Las políticas de Gestión Pública (MIPG) y de Gobierno Digital orientan la simplificación de trámites y la gestión del conocimiento como pilares para mejorar la oportunidad y calidad de la respuesta institucional, especialmente cuando se interpretan múltiples fuentes como normatividad, lineamientos, manuales y datos operativos (DAFP, 2023; MinTIC, 2022).

Normatividad que usa el agente

El agente opera sobre un corpus normativo conformado por los instrumentos que regulan el Proceso de Selección DIAN 2676, convocado por la CNSC mediante el Acuerdo No. 21 del 7 de noviembre de 2025. Este corpus está integrado por dos documentos: el Acuerdo de convocatoria, que establece las reglas generales del proceso, incluyendo la estructura del concurso, los requisitos generales de participación, las pruebas a aplicar y su ponderación, los criterios de conformación de listas de elegibles y las causales de exclusión; y el Anexo Técnico, que desarrolla las especificaciones técnicas de cada fase, entre ellas los procedimientos de inscripción en el SIMO, las definiciones y condiciones para la verificación de requisitos mínimos, los criterios valorativos de educación y experiencia para la Prueba de Valoración de Antecedentes, y las reglas de aplicación de las Pruebas Escritas (CNSC, 2025a, 2025b).

Ambos documentos se indexan en el repositorio del agente con metadatos de fuente, fecha de expedición y numeral, de modo que cada fragmento recuperado es trazable hasta su origen. Este diseño garantiza que los borradores generados se fundamenten en referencias verificables, en consonancia con los principios de motivación administrativa y trazabilidad exigidos por la Ley 1755 de 2015 (Congreso de la República, 2015).

Recuperación de información

Desde el punto de vista de la gestión de la información, la recuperación eficiente de normas, precedentes y criterios internos es una condición importante para fundamentar respuestas coherentes con la regulación vigente. La recuperación de información proporciona los conceptos de consulta, pertinencia e índice que permiten ubicar pasajes relevantes en

grandes repositorios textuales, así como los fundamentos de evaluación como precisión y cobertura para asegurar utilidad práctica en entornos institucionales (Manning, Raghavan y Schütze, 2008).

La escalabilidad de la recuperación depende de índices y estructuras vectoriales que soporten consultas semánticas sobre colecciones jurídicas en crecimiento. Bibliotecas y métodos de búsqueda aproximada de vecinos más cercanos posibilitan tiempos de respuesta bajos sin sacrificar pertinencia, especialmente cuando se complementan con metadatos de vigencia, fuente y jerarquía normativa para asegurar que los resultados sean válidos en el tiempo (Johnson, Douze y Jégou, 2019; Manning et al., 2008).

Procesamiento de Lenguaje Natural

El Procesamiento de Lenguaje Natural (PLN) aporta técnicas para identificar en los documentos administrativos componentes clave como hechos, pretensiones, asuntos, anexos y referencias normativas, reduciendo de esta manera la variabilidad textual y facilitando los cruces automáticos con fuentes operativas. Técnicas como el etiquetado morfosintáctico y el reconocimiento de entidades permiten marcar cada palabra con su función gramatical e identificar nombres y datos relevantes, mientras que la segmentación por patrones estructura el contenido y habilita validaciones de negocio antes de la redacción final, incluso cuando el material proviene de formularios o escritos no estandarizados (Jurafsky y Martin, s. f.).

Desde un punto de vista tecnológico, los avances en PLN y en Procesamiento Inteligente de Documentos (IDP) han logrado clasificar, extraer y validar información en expedientes y comunicaciones oficiales, reduciendo cargas manuales y variabilidad de criterios (Jiang et al., 2023; OCDE, 2025). Por el lado del dominio legal, las tareas especializadas como la respuesta a preguntas con soportes documentales y la segmentación de apartados han madurado lo suficiente para operar en flujos reales de trabajo, siempre que se controle la vigencia normativa y la trazabilidad de las citas (Lai et al., 2024; Ariai et al., 2024).

Generación Aumentada por Recuperación (RAG)

Para fijar la redacción en evidencias y reducir errores, las arquitecturas de Generación Aumentada por Recuperación (RAG) combinan un componente de búsqueda que recupera pasajes pertinentes desde un corpus con un generador que redacta únicamente a partir de esos pasajes, conservando las citas de origen. En dominios normativos, este enfoque favorece la fidelidad factual y la verificación, dado que cada afirmación puede vincularse al fragmento

normativo que la respalda, preservando la cadena de custodia argumentativa (Lewis et al., 2020).

Investigaciones recientes advierten sobre el riesgo de errores de recuperación y alucinaciones cuando los modelos generativos producen textos sin anclarlos a fuentes verificables. Para mitigar esta limitación, la arquitectura RAG se ha consolidado como un enfoque de referencia, al combinar la recuperación previa de pasajes relevantes desde un corpus versionado con la generación de texto respaldado por referencias explícitas (Klesel y Wittmann, 2025). En el ámbito jurídico, evaluaciones como LegalBench-RAG evidencian que no basta con recuperar documentos completos, sino que la precisión en la recuperación a nivel de fragmentos es determinante para construir respuestas normativas sólidas (Pipitone y Houir, 2024). A nivel de infraestructura, bibliotecas de búsqueda vectorial como FAISS permiten crear índices eficientes y reproducibles, adecuados para gestionar colecciones normativas en constante crecimiento (Johnson et al., 2019).

Fundamentación matemática de la búsqueda híbrida y de las arquitecturas RAG

La pertinencia de la arquitectura adoptada se sustenta en tres cuerpos formales complementarios: la búsqueda híbrida (densa y dispersa) con fusión de rankings, la arquitectura Transformer que habilita los modelos de embeddings y de lenguaje, y las métricas específicas para evaluar sistemas RAG. A continuación, se desarrollan de forma sintética, en línea con la implementación efectiva del prototipo.

Búsqueda densa, búsqueda dispersa y su complementariedad

La recuperación de información combina dos paradigmas que se complementan. La búsqueda densa, de carácter semántico, representa la consulta y los fragmentos como vectores en un espacio de embeddings y mide su afinidad mediante la similitud del coseno, definida como:

$$\text{sim}(q, d) = (q \cdot d) / (||q|| \cdot ||d||)$$

Donde q y d son los vectores de la consulta y del fragmento. Cuando los vectores se normalizan a norma unitaria, la similitud del coseno equivale al producto interno, propiedad que aprovecha el índice FAISS de tipo IndexFlatIP del prototipo. Esta búsqueda captura relaciones de significado entre términos jurídicos equivalentes, pero puede pasar por alto coincidencias exactas de numerales, artículos o nombres de acuerdos. Para corregir esa limitación, la

búsqueda dispersa, de carácter léxico, pondera la coincidencia exacta de términos mediante la función BM25:

$$BM25(q, d) = \sum IDF(t_i) \cdot [f(t_i, d) \cdot (k_1 + 1)] / [f(t_i, d) + k_1 \cdot (1 - b + b \cdot |d| / avgdl)]$$

donde $f(t_i, d)$ es la frecuencia del término t_i en el fragmento d , $|d|$ su longitud, $avgdl$ la longitud media de los fragmentos e $IDF(t_i)$ el peso de frecuencia inversa de documento. El prototipo emplea los valores estándar $k_1 = 1,5$ y $b = 0,75$. La combinación de ambas búsquedas resulta natural en el dominio de la CNSC, donde una consulta puede requerir tanto comprensión semántica como precisión literal en la referencia normativa.

Para integrar los dos rankings sin depender de la escala de sus puntajes se aplica la Fusión Recíproca de Rangos (Reciprocal Rank Fusion, RRF):

$$RRF(d) = \sum 1 / (k + r_i(d))$$

Donde $r_i(d)$ es la posición o rango del fragmento d en el sistema i (denso o disperso) y k es una constante de amortiguación fijada en 60. La RRF favorece a los fragmentos bien posicionados en ambos sistemas y es robusta frente a diferencias de magnitud entre puntajes. Sobre el conjunto fusionado, un modelo de tipo Cross-Encoder (el reranker mmarco-mMiniLMv2-L12-H384-v1) recalifica cada par consulta-fragmento procesándolos de forma conjunta, lo que produce una estimación de relevancia más precisa que la comparación de vectores independientes y mitiga los falsos positivos. Esta mitigación es especialmente valiosa en el derecho administrativo, donde un fragmento puede ser superficialmente similar pero jurídicamente inaplicable. En la configuración del prototipo, el puntaje final pondera la señal del reranker (0,75) y el puntaje híbrido fusionado (0,25), seguido de una selección diversificada que limita la redundancia y el número de fragmentos por documento fuente.

Arquitectura Transformer y modelos de lenguaje pequeños (SLM)

Tanto el modelo de embeddings como el modelo generativo se fundamentan en la arquitectura Transformer (Vaswani et al., 2017), cuyo mecanismo central es la auto-atención, definida como:

$$Attention(Q, K, V) = softmax((Q \cdot K^T) / \sqrt{d_k}) \cdot V$$

Donde Q , K y V son las matrices de consultas, claves y valores, y d_k la dimensión de las claves. Este mecanismo pondera la relevancia relativa de cada token frente a los demás y captura dependencias de largo alcance, base del procesamiento contextual del lenguaje. El auge de los modelos de lenguaje pequeños (Small Language Models, SLM), del orden de mil a siete mil millones de parámetros, como Qwen2.5 de tres mil millones (3B), responde a la búsqueda de un equilibrio entre capacidad y costo de cómputo. Estos modelos ofrecen un desempeño suficiente para tareas acotadas y guiadas por contexto, como la generación restringida a la evidencia recuperada, con requisitos de memoria e inferencia que los hacen viables en infraestructura modesta e incluso sin GPU, lo que habilita el despliegue local y la soberanía del dato. Su contrapartida es una menor capacidad de razonamiento abstracto, que en el prototipo se compensa con ingeniería de prompts y anclaje estricto a la evidencia (grounding), restringiendo la generación a los fragmentos normativos efectivamente recuperados.

Métricas de evaluación de arquitecturas RAG

La evaluación de un sistema RAG distingue la calidad de la recuperación de la calidad de la generación. En el componente de recuperación se emplean métricas de relevancia y precisión de contexto, como la precisión en los primeros k resultados recuperados y la exhaustividad o cobertura correspondiente, que miden, respectivamente, la proporción de fragmentos pertinentes entre los k recuperados y la cobertura de los fragmentos pertinentes existentes. En el dominio jurídico, evaluaciones como LegalBench-RAG muestran que la precisión a nivel de fragmento, y no solo a nivel de documento, es determinante (Pipitone y Houir Alami, 2024). En el componente de generación, la fidelidad evalúa que cada afirmación del borrador esté respaldada por la evidencia recuperada, y la relevancia de la respuesta evalúa su pertinencia frente a la solicitud. Estas dimensiones fundamentan tanto el diseño del evaluador automático de calidad del prototipo, que verifica la presencia y la trazabilidad de las citas, como la interpretación de los hallazgos del piloto sobre la precisión y la completitud de las citas normativas.

Comparación de enfoques

Los enfoques revisados presentan ventajas y límites diferenciados en contextos públicos. RAG mejora la fidelidad factual al anclar la redacción en pasajes recuperados, aunque requiere un corpus versionado y puede degradarse si hay desalineación de vigencias. El enfoque híbrido de PLN con reglas y modelos para segmentar la solicitud del aspirante

reduce la variabilidad textual y favorece la trazabilidad, pero presenta el riesgo de desfase entre los formatos utilizados por los ciudadanos y la necesidad de contar con un dataset anotado para recalibrar. En lo sociotécnico, una implementación de IA sin explicabilidad y sin revisión humana incrementa el riesgo de citas descontextualizadas y conflictos con el debido proceso.

Esta comparación justifica la incorporación de controles de vigencia y jerarquía normativa, registros de fuentes, umbrales de confianza y revisión jurídica obligatoria antes del envío de cualquier respuesta.

Gobernanza de datos y seguridad de la información

La gobernanza de datos y la seguridad de la información son condiciones habilitantes para un uso responsable de IA en el sector público. Marcos como ISO/IEC 27001 guían controles sobre confidencialidad, integridad y disponibilidad (ISO/IEC, 2022). Por otro lado, DAMA-DMBOK orienta la gestión del ciclo de vida de los datos, incluyendo los fundamentos, buenas prácticas y funciones del análisis de datos (DAMA International, 2017). Los principios de IA confiable promueven trazabilidad y gestión de riesgo durante todo el ciclo de desarrollo y uso. Integrar bitácoras, control de versiones normativas, plantillas y validación de vigencias contribuye a decisiones auditables y a la rendición de cuentas (NIST, 2023; OCDE, 2019).

Componente sociotécnico

El componente sociotécnico es igualmente central en el diseño de la solución. Los sistemas de apoyo a decisiones públicas deben incorporar revisión humana informada, explicaciones útiles y controles para la intervención del funcionario cuando existan ambigüedades o señales de baja confianza. Las guías de interacción entre humanos e IA enfatizan la necesidad de exponer supuestos, evidencias utilizadas y límites de la recomendación, reforzando la capacidad del equipo para detectar errores y ajustar criterios (Amershi et al., 2019; Doshi-Velez y Kim, 2017). Estudios acerca de la adopción responsable de IA en gobiernos recomiendan incorporar controles de confianza, como versionado de normas, reglas de vigencia y umbrales de confianza en la recuperación, prácticas coherentes con los principios de legalidad, igualdad y debido proceso en la administración pública (Agbabiaka et al., 2025; Neumann et al., 2022).

Antecedentes institucionales y técnico-organizacionales

A nivel institucional, la digitalización del Estado en países hispanohablantes ha avanzado desde la automatización de trámites hacia esquemas de gestión documental

electrónica con metadatos normalizados, control de versiones y trazabilidad de decisiones (ISO 15489-1, 2016). En este tránsito, la estandarización de procesos documentales, la gobernanza de datos y la analítica aplicada al texto han mostrado impactos positivos en oportunidad y consistencia, sobre todo cuando la evidencia se organiza como conocimiento reutilizable para el servidor público (AGN, 2024). Estos esfuerzos se articulan con políticas de gestión pública y gobierno digital que promueven interoperabilidad, seguridad de la información y rendición de cuentas, reduciendo variabilidad en la respuesta y fortaleciendo la trazabilidad administrativa (MinTIC, 2022).

En el plano técnico-organizacional, los repositorios normativos y de criterios internos han evolucionado desde buscadores por palabra clave hacia mecanismos de recuperación semántica que contemplan vigencias, jerarquías y relaciones entre normas, manuales y lineamientos. A su vez, la segmentación de documentos administrativos en componentes como asunto, hechos, pretensión, anexos y citas facilita cruces automáticos con fuentes operativas y habilita verificaciones previas a la redacción. La literatura sectorial coincide en que estos elementos, la estandarización, la gobernanza y la recuperación de conocimiento, operan de forma sinérgica, dado que ninguno por sí solo asegura oportunidad ni consistencia institucional (AGN, 2001; Senado de la República, 2022).

Brechas identificadas

Pese a estos avances, persisten brechas específicas para el servidor público que debe responder peticiones: dispersión de fuentes (normatividad, instructivos y precedentes en repositorios separados), heterogeneidad de formatos provenientes de los ciudadanos, dificultad para verificar vigencias y numerales al momento de citar, y ausencia de flujos que expongan claramente la evidencia utilizada y los límites de la recomendación automatizada. Estas brechas se acentúan en etapas de alto volumen, donde el cuello de botella no es solo recuperar normas, sino articularlas con los datos del caso y dejar trazas auditables que sustenten la decisión. De allí la pertinencia de un asistente especializado que integre recuperación guiada por vigencias, extracción estructurada y redacción apoyada en evidencias, manteniendo revisión humana antes del envío y alineación con las políticas institucionales (OCDE, 2019; CNSC, s. f.).

Articulación entre referentes teóricos y la solución propuesta

En conjunto, el marco normativo (Ley 1755 de 2015; Ley 909 de 2004; Decreto 1083 de 2015) y los lineamientos de gobierno de datos y gestión documental en el Estado, junto con los

avances en recuperación de información, PLN y RAG, convergen en una misma necesidad: responder peticiones con oportunidad, fundamento verificable y trazabilidad (Congreso de la República, 2004, 2015; Presidencia de la República, 2015; DAMA International, 2017; ISO/IEC, 2022; DAFP, 2023; MinTIC, 2019, 2022). Esta convergencia sustenta la pertinencia de un asistente interno que integre extracción estructurada de la solicitud del aspirante, cruce controlado con SIMO/CSV bajo minimización de datos y borradores anclados en fragmentos normativos vigentes citados con precisión, manteniendo revisión humana informada (Manning et al., 2008).

La literatura converge en un enfoque híbrido que aborda los vacíos detectados: corpus normativo versionado con control de vigencias y numerales, PLN con reglas para identificar y segmentar la solicitud del aspirante con trazabilidad, RAG con recuperación guiada por vigencia y jerarquía con citas exactas, y cruce SIMO/CSV con registros de auditoría. Todo ello bajo principios de IA confiable, gestión de riesgo, seguridad de la información y revisión humana, orientado a la oportunidad, claridad y motivación suficiente de la respuesta (Lewis et al., 2020; Jurafsky y Martin, s. f.).

En síntesis, la literatura y los marcos de referencia convergen en que la combinación de recuperación de información, PLN y RAG bajo gobernanza y revisión humana es pertinente para entornos normativos exigentes como la CNSC. Persisten vacíos específicos que este proyecto aborda: (1) evaluación de recuperación a nivel de fragmento para normatividad colombiana en español; (2) integración práctica entre datos operativos (SIMO/CSV) y repositorios normativos versionados; (3) mecanismos de explicación de citas útiles para la revisión jurídica; y (4) trazabilidad desde la evidencia hasta la decisión. Con ello, el marco de referencia respalda que un asistente interno puede contribuir a reducir tiempos de elaboración y a aumentar la exactitud de cita, aportando valor público sin comprometer legalidad ni transparencia.

Para efectos de síntesis, la articulación entre los referentes teóricos y la solución propuesta se organiza en decisiones de diseño, controles verificables e indicadores de desempeño.

Concepto teórico	Decisión de diseño	Control / Práctica
Recuperación de información y pertinencia	RAG estricta: recuperar pasajes y generar solo a partir de ellos	Índice versionado con vigencias y jerarquía; logs de fuentes
PLN y segmentación estructural	Segmentador de peticiones multi-pregunta con detección de conectores	Checklist de límites, numeración y trazabilidad

Gobernanza de datos y seguridad (ISO 27001, DAMA)	Bitácoras, versionado, perfiles de acceso y minimización	Registro de decisiones, auditoría y control de cambios
IA confiable y revisión humana (NIST AI RMF)	Umbral de confianza, evaluador A/B/C y revisión jurídica obligatoria	Alertas por baja confianza; explicación de citas
MIPG y Gobierno Digital	Plantillas institucionales y metadatos para trazabilidad	Campos obligatorios: norma, artículo, numeral y vigencia

Tabla 1 Vínculos entre conceptos teóricos y componentes de la solución propuesta. Elaboración propia a partir de los referentes teóricos consultados.

Diseño Metodológico

Tipo de investigación

La presente investigación es aplicada, orientada al diseño y validación de un asistente interno de respuestas a peticiones en procesos de selección por mérito de la Comisión Nacional del Servicio Civil (CNSC). Su propósito es transformar el conocimiento teórico en soluciones prácticas para optimizar los tiempos de respuesta, la trazabilidad y la precisión normativa en la atención ciudadana.

Según Hernández Sampieri, Fernández y Baptista (2014), la investigación aplicada busca resolver problemas reales mediante el uso de teorías, métodos y técnicas científicas. El alcance es exploratorio-descriptivo, dado que pretende caracterizar el proceso actual y generar una primera aproximación al desempeño del prototipo, sin pretender aún establecer causalidades.

El enfoque de los datos es mixto. Se combinará análisis cualitativo de normativa y documentos (revisión y codificación de requerimientos) con métricas cuantitativas de desempeño del prototipo. La inferencia será principalmente deductiva, partiendo del marco normativo para derivar reglas y criterios de validación, con ajustes iterativos cuando la evidencia lo sugiera.

Análisis externo

Para el presente proyecto se adopta la metodología DOFA, con el propósito de identificar las fortalezas, oportunidades, debilidades y amenazas que inciden en el desarrollo e implementación del asistente de inteligencia artificial orientado a la generación de borradores de respuesta (Chiavenato y Sapiro, 2019). Este análisis permite comprender el entorno institucional y tecnológico en el que se enmarca el proyecto, facilitando la definición de estrategias que aprovechen las ventajas existentes y mitiguen los riesgos potenciales.

La matriz DOFA se fundamenta en el diagnóstico interno y externo de la entidad, considerando tanto los factores de contexto político-normativo como las condiciones técnicas y operativas actuales. La Figura 4 presenta la matriz DOFA elaborada para el proyecto.

	Aspectos negativos	Aspectos positivos
Análisis interno	Debilidades - Dependencia de la calidad y actualización del corpus normativo y documental para entrenar el modelo. - Infraestructura tecnológica limitada para pruebas locales en algunos entornos institucionales. - Brecha inicial de capacidades en inteligencia artificial aplicada a procesos administrativos. - Falta de protocolos estandarizados para medir precisión, fidelidad y trazabilidad del texto generado.	Fortalezas - Existencia de un marco normativo consolidado que regula el trámite de peticiones (Ley 909/2004, Decreto 1083/2015, Ley 1755/2015). - Disponibilidad de plantillas institucionales y repositorios normativos que facilitan la estructuración de respuestas. - Conocimiento técnico y experiencia del equipo en el manejo de peticiones y en los procesos de inscripciones. - Licencias institucionales como Gemini Pro y recursos tecnológicos que permiten desarrollar el asistente sin incurrir en costos adicionales.
Análisis externo	Amenazas - Riesgos de seguridad y confidencialidad en el tratamiento de datos personales de aspirantes. - Cambios normativos o directrices institucionales que puedan alterar los procedimientos o plazos de respuesta. - Resistencia al cambio y percepción de sustitución de funciones por parte de algunos funcionarios. - Riesgo reputacional si el sistema no comunica claramente su rol como herramienta de apoyo y no como emisor final de decisiones.	Oportunidades - Impulso del Gobierno Digital y las políticas de Transformación Digital del Estado, que promueven la adopción de IA en la gestión pública. - Avances tecnológicos en procesamiento de lenguaje natural (PLN) y recuperación semántica, que posibilitan soluciones locales y seguras. - Demanda ciudadana creciente por respuestas más ágiles, trazables y ajustadas a la norma. - Potencial de escalar el asistente a otras etapas del proceso de selección una vez validado el piloto de inscripciones.

Figura 4 Matriz DOFA, elaboración propia

Del análisis anterior se infiere que el entorno externo es favorable para la implementación del asistente, especialmente por el impulso normativo y tecnológico hacia la digitalización del servicio público. Sin embargo, la sostenibilidad del proyecto dependerá de la gestión de riesgos asociados a la seguridad de datos y la adopción institucional. Las estrategias recomendadas se orientan a potenciar las fortalezas internas mediante la estructuración de un repositorio normativo controlado y actualizable, aprovechar las oportunidades externas a través del alineamiento con políticas nacionales de innovación pública, mitigar debilidades mediante capacitaciones técnicas y validación incremental del prototipo, y prevenir amenazas fortaleciendo los protocolos de privacidad, control de versiones y comunicación interna sobre el alcance del asistente.

Análisis interno

El análisis interno se desarrolló mediante la aplicación de una encuesta tipo Likert diseñada para evaluar la percepción de los funcionarios sobre el desempeño del asistente de inteligencia artificial en el proceso de gestión de peticiones de la CNSC. De acuerdo con Bernal

(2020), los modelos diagnósticos permiten identificar las fortalezas, debilidades y brechas operativas que inciden en la eficiencia y calidad de las respuestas institucionales, con el fin de establecer la línea base sobre la cual se desarrolla y mejora el asistente inteligente propuesto.

El instrumento se estructuró en siete dimensiones: (1) apoyo normativo, que evalúa la pertinencia y precisión de las referencias legales sugeridas por el agente; (2) calidad de las citas, que mide la exactitud y vigencia de los artículos y normas recuperados; (3) eficiencia, que recoge la percepción sobre los tiempos de respuesta y el aporte del agente a la reducción de la carga de trabajo; (4) calidad de la redacción, que valora la coherencia, el lenguaje institucional y la estructura de los borradores generados; (5) comprensión de la solicitud, que verifica si el agente identifica correctamente el objeto de la petición y da una respuesta de fondo; (6) experiencia de usuario, que evalúa la usabilidad, claridad de la interfaz y facilidad de adopción de la herramienta; y (7) confianza y adopción, que mide la disposición de los funcionarios a incorporar el asistente en su flujo de trabajo habitual.

La encuesta fue aplicada durante los días 27 y 28 de abril de 2026 a cinco (5) funcionarios vinculados al proceso de atención de peticiones del Proceso de Selección DIAN 2676, constituyendo una muestra censal de los participantes disponibles en las sesiones del piloto. Adicionalmente, se aplicó un cuestionario de percepción con preguntas abiertas a dos (2) participantes, con el fin de contextualizar y triangular los hallazgos cuantitativos. Los datos obtenidos se sistematizaron y analizaron mediante estadística descriptiva, calculando medias por ítem y por dimensión, así como el Alfa de Cronbach como indicador de consistencia interna del instrumento.

Población, muestra y ficha técnica

La población objeto del estudio corresponde al personal vinculado al proceso de selección que atiende peticiones en la CNSC, incluyendo funcionarios responsables de la elaboración de respuestas, revisores jurídicos, asesores de proceso de selección y personal de apoyo técnico. La Tabla 2 presenta la ficha técnica de la investigación.

Elemento	Descripción
Tipo de investigación	Aplicada, de enfoque mixto, con alcance exploratorio-descriptivo
Unidad de análisis	Funcionarios de la CNSC vinculados a la gestión de peticiones del Proceso de Selección DIAN 2676

Población	9 personas directamente vinculadas al proceso de atención de peticiones del Proceso de Selección DIAN: 5 funcionarios elaboradores, 1 líder jurídico, 1 asesor del proceso, 2 apoyos técnicos
Muestra	Cinco (5) funcionarios para la encuesta Likert; dos (2) participantes para el cuestionario de preguntas abiertas (cuestionario de percepción)
Instrumentos	Encuesta tipo Likert de 25 ítems estructurada en siete dimensiones y cuestionario de percepción con 13 preguntas abiertas (información general más experiencia, eficiencia, calidad, barreras, recomendaciones)
Técnicas de análisis	Estadística descriptiva (medias, desviaciones estándar, distribuciones de frecuencia), Alfa de Cronbach por dimensión, análisis categorial de las respuestas abiertas y triangulación de hallazgos
Periodo de recolección	Fase de implementación del piloto controlado (13 a 30 de abril)
Confidencialidad	Anonimización de datos, consentimiento informado y uso exclusivo para fines académicos

Tabla 2 Ficha técnica de la investigación. *Elaboración propia*

Para la aplicación de los instrumentos se emplearon dos técnicas complementarias: una encuesta estructurada tipo Likert (escala de 1 a 5) con 25 ítems distribuidos en siete dimensiones de evaluación, y un cuestionario de preguntas abiertas con 13 preguntas distribuidas en seis bloques temáticos. La encuesta Likert fue diligenciada por cinco funcionarios y el cuestionario de preguntas abiertas por dos funcionarios. La diferencia entre ambos tamaños obedece a la disponibilidad de los participantes en la ventana de aplicación, dado que la participación fue voluntaria; los hallazgos cualitativos del cuestionario se contrastaron con los resultados cuantitativos de la encuesta para triangular ambas fuentes.

La población está conformada por nueve (9) personas directamente vinculadas al proceso de atención de peticiones del Proceso de Selección DIAN. Incluye cinco (5) funcionarios responsables de la elaboración de respuestas, un (1) líder jurídico encargado de la revisión, un (1) asesor del proceso de selección que firma la respuesta para su envío, y dos (2) apoyos técnicos a cargo de la asignación y trazabilidad en el sistema. Se empleará muestreo censal; por tanto, la muestra coincide con la población definida.

Identificación de variables

La variable independiente corresponde a la implementación del modelo diagnóstico de madurez del proceso de respuestas a peticiones, entendida como el conjunto de acciones metodológicas y técnicas orientadas a evaluar las capacidades normativas, tecnológicas, operativas y humanas de la CNSC (Arias, 2012). La variable dependiente es la eficiencia y calidad del proceso de respuesta a peticiones, que representa el nivel de desempeño institucional medido a partir de indicadores de estandarización, precisión normativa, tiempos de respuesta y satisfacción del funcionario con el proceso (Arias, 2012).

Instrumento de medición

Se aplicaron dos instrumentos. El primero es una encuesta estructurada tipo Likert con 25 ítems valorados en una escala de 1 a 5 (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo), distribuidos en siete dimensiones de evaluación: (D1) apoyo en la identificación normativa y estructuración de la respuesta, con 4 ítems; (D2) calidad y precisión de las citas normativas, con 4 ítems; (D3) eficiencia y reducción de tiempos, con 3 ítems; (D4) calidad de redacción y estructura argumentativa, con 3 ítems; (D5) comprensión e interpretación de la petición, con 3 ítems; (D6) experiencia de usuario e interfaz, con 4 ítems; y (D7) confianza, revisión humana y disposición de adopción, con 4 ítems. El segundo instrumento es un cuestionario de preguntas abiertas con 13 preguntas (más datos generales) que indaga sobre la primera experiencia con el agente, la utilidad del borrador, la calidad de las citas, el manejo de peticiones multi-pregunta, el impacto en eficiencia y calidad, las limitaciones percibidas, las funcionalidades deseadas, las barreras de adopción y la valoración general.

Validación del instrumento

La validación del instrumento Likert se desarrolló en dos fases. En la primera fase (validez de contenido), las 25 afirmaciones fueron revisadas en términos de claridad y pertinencia por dos revisores: el director del trabajo de grado y un funcionario con experiencia en el proceso de atención de peticiones. Cada revisor verificó la cobertura de las siete dimensiones definidas, la pertinencia de cada ítem respecto de la dimensión que pretende medir y la comprensibilidad del lenguaje; las observaciones se registraron y dieron lugar a ajustes de redacción en los ítems señalados como ambiguos antes de la aplicación. En la segunda fase (confiabilidad interna), una vez aplicado el instrumento se calculó el coeficiente Alfa de Cronbach a nivel global y por dimensión, con un umbral de referencia de 0,70. Los valores obtenidos se reportan en la sección de resultados; conforme a lo allí señalado, la interpretación psicométrica se restringe al Alfa global dado el tamaño muestral reducido ($N = 5$).

Se reconoce como limitación de validación que el procedimiento se apoyó en dos revisores y no en un panel de jueces expertos. La práctica recomendada para instrumentos psicométricos aplicados en contextos institucionales contempla un panel de tres a cinco jueces con cuantificación del acuerdo inter-jueces mediante coeficientes como la V de Aiken o el Kappa de Fleiss; este cálculo no se realizó en el presente estudio. Esta limitación de validez de contenido, sumada al tamaño muestral reducido, refuerza el carácter exploratorio del diagnóstico y debe tenerse presente al interpretar las dimensiones D1 a D7. En aplicaciones futuras del instrumento se recomienda incorporar el panel de jueces y el cálculo de la V de Aiken como paso previo a la aplicación.

Para el cuestionario de percepción de preguntas abiertas, la validación se limitó a la revisión de contenido por parte del director del trabajo de grado, quien verificó la pertinencia temática de los seis bloques y la claridad del lenguaje. Al tratarse de un instrumento cualitativo complementario, no se calcularon indicadores de confiabilidad estadística; su rigor se sustenta en la coherencia entre las preguntas y los objetivos del estudio, y en la triangulación de sus hallazgos con los resultados cuantitativos de la encuesta Likert.

Diagnóstico Organizacional

El diagnóstico organizacional del proceso de atención de peticiones en la CNSC se construyó a partir de tres fuentes complementarias: el análisis DOFA presentado en el diseño metodológico, la línea base operativa derivada de métricas internas del proceso de selección y la revisión del flujo de trabajo actual mediante observación directa del proceso.

Línea base operativa

En cuanto a la línea base, se tomaron como referencia los datos del reporte de control de términos registrados entre el 20 de agosto y el 20 de octubre de 2025 para el Grupo de Gerencia de Convocatoria 12 del Proceso de Selección DIAN 2676. El reporte contiene 416 registros de gestión que corresponden a 328 peticiones únicas; los registros adicionales se originan cuando un mismo radicado genera más de una respuesta, por ejemplo en los traslados a otras entidades o a otras áreas internas. Sobre las 328 peticiones con ciclo completo registrado, el tiempo promedio transcurrido desde la radicación hasta la comunicación al peticionario fue de 6,5 días hábiles. Al desagregar el ciclo por hitos del proceso se obtuvieron los resultados que se presentan en la siguiente tabla.

Hito del proceso	Tiempo promedio (días hábiles)	Registros considerados (n)
Asignación de la petición al responsable	0,38	416
Elaboración del borrador (asignación a revisión)	3,41	356
Revisión jurídica	2,53	254
Firma y envío	0,98	250
Tiempo total promedio (radicación a contestación)	6,55	328

Tabla 3 *Tiempos promedio por hito del proceso de atención de peticiones (periodo agosto-octubre 2025). Elaboración propia a partir de datos tomados de la CNSC (2025c).*

Estos datos confirman que el tramo más exigente del ciclo es la elaboración de la respuesta una vez la petición llega al funcionario responsable, con un promedio de 3,41 días hábiles, equivalente a poco más de la mitad del tiempo total promedio del ciclo (6,55 días). Este hallazgo es consistente con un trabajo intensivo de análisis del caso, verificación de antecedentes en múltiples fuentes y construcción de argumentación jurídica con citas normativas precisas.

Nota metodológica. La tabla reporta la duración promedio de cada hito, calculada sobre los registros que cuentan con la fecha de ese hito (columna n); no constituye una partición del tiempo total y, por ello, los promedios por hito no suman el total. El tiempo total promedio (6,55 días hábiles) se mide de forma independiente, de extremo a extremo, sobre las 328 peticiones con ciclo completo. El reporte contiene 416 registros de gestión que corresponden a 328 peticiones únicas; los registros adicionales provienen de radicados que generan más de una respuesta, como los traslados a otras entidades o áreas. Los valores coinciden con los indicadores oficiales de la entidad (CNSC, 2025c) y reemplazan los reportados en versiones previas del documento.

La Figura 5 presenta el mapa del proceso actual para la elaboración de respuestas a peticiones en la CNSC, mostrando los actores involucrados en cada etapa y los puntos de decisión.

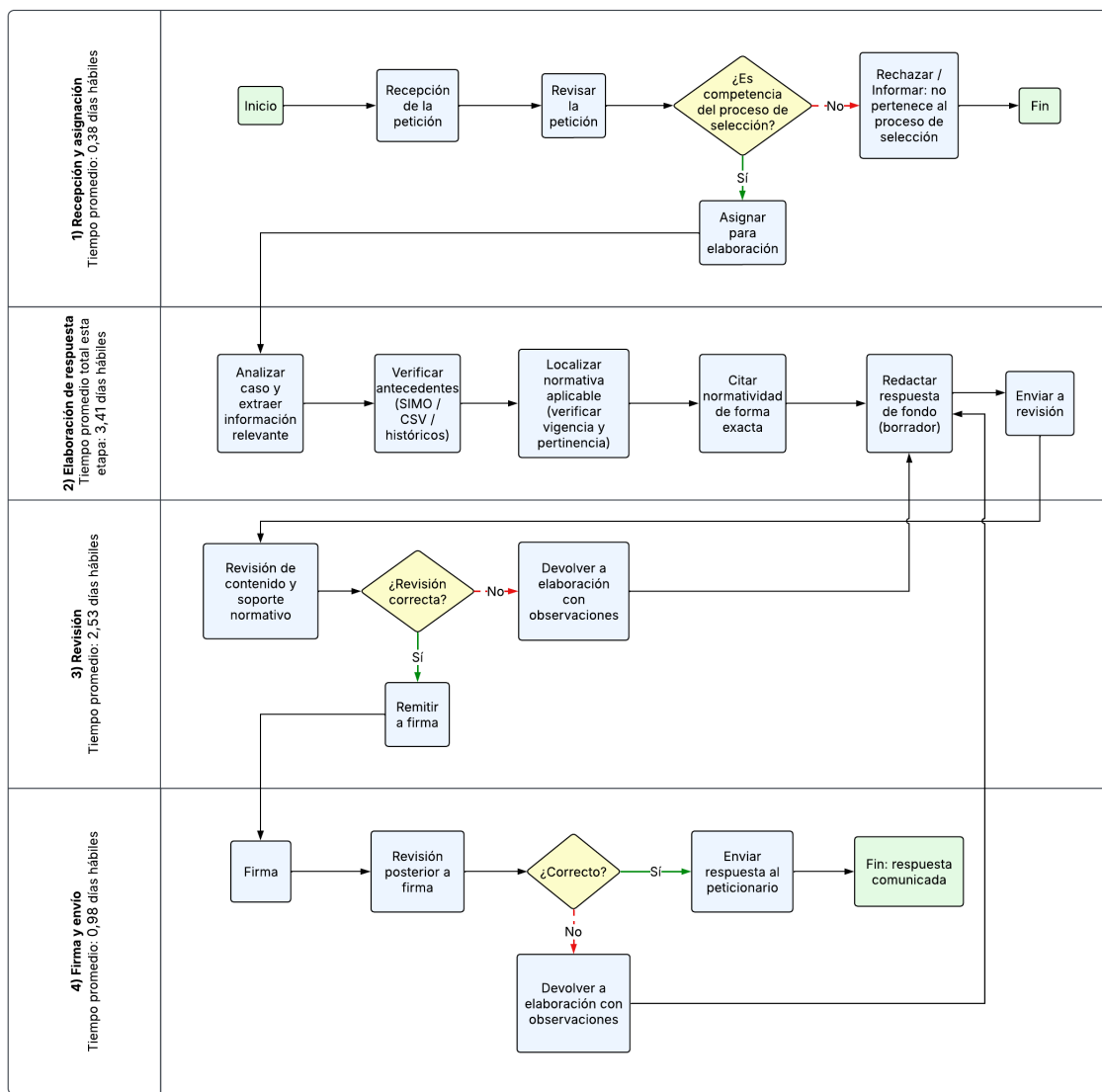


Figura 5 Mapa del proceso actual para la elaboración de respuestas a peticiones en la CNSC. Elaboración propia.

La Figura 6 muestra el proceso detallado desde la creación de la petición hasta la respuesta al peticionario.

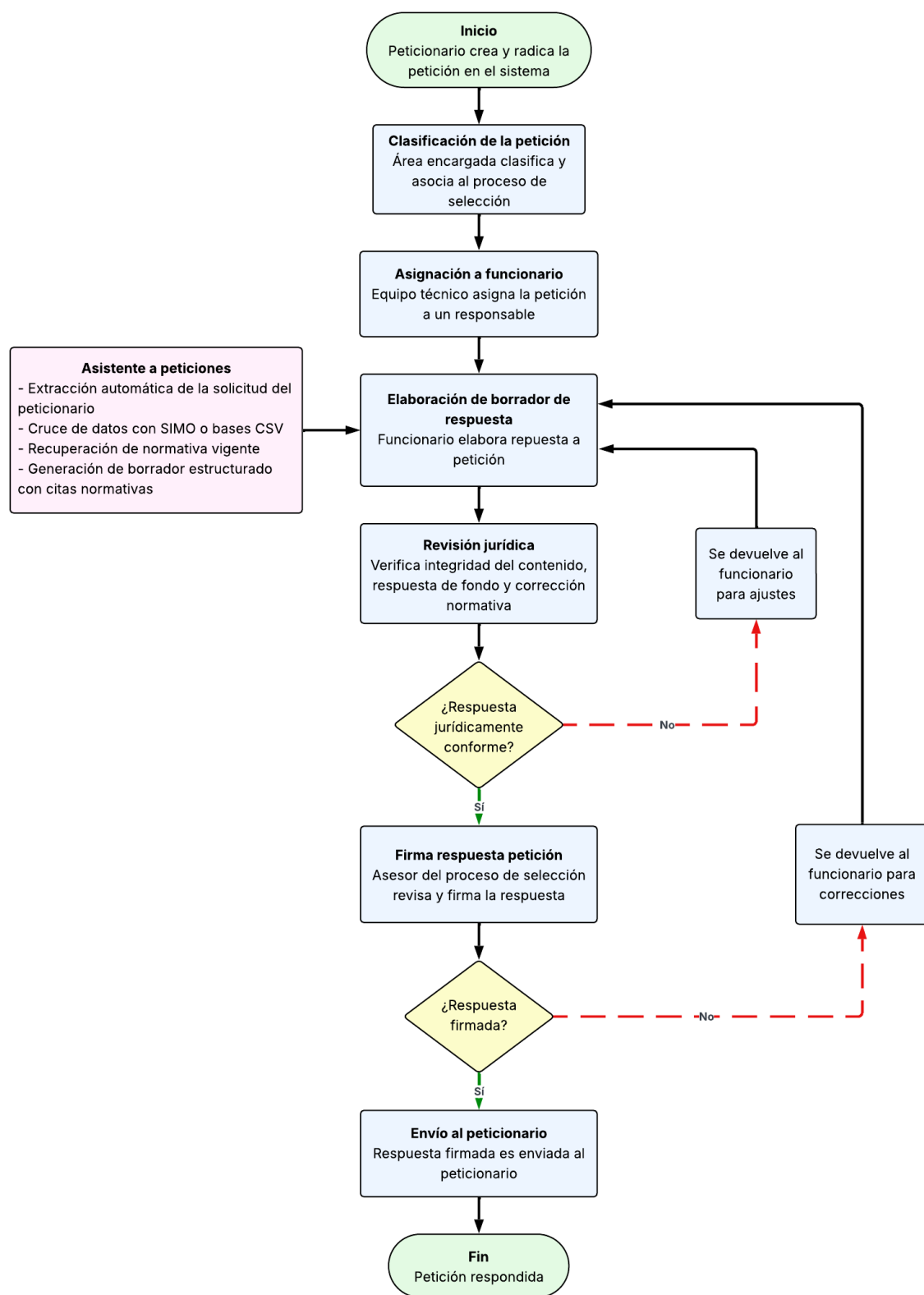


Figura 6 Proceso desde la creación de la petición hasta la respuesta al peticionario, con tiempos promedio por hito. Elaboración propia a partir de datos tomados de la CNSC.

Situación actual

El flujo actual de atención de peticiones opera de la siguiente manera. Una vez la petición es recibida y radicada por el área de Atención al Ciudadano, se verifica la competencia de la CNSC para responder y se clasifica por etapa del proceso de selección. Posteriormente, la petición se asigna al funcionario responsable, quien debe leer el texto completo para identificar las solicitudes del ciudadano, consultar los datos del aspirante en la plataforma SIMO o en archivos CSV, localizar la normatividad aplicable entre múltiples fuentes (acuerdos de convocatoria, anexos, decretos, leyes y lineamientos) y redactar el borrador de respuesta con fundamentación jurídica. Una vez elaborado, el borrador pasa a revisión jurídica y, si es aprobado, a firma y envío. Si el borrador presenta deficiencias en la argumentación, citas incorrectas o ausencia de respuesta de fondo, se devuelve al funcionario para corrección, lo que incrementa el tiempo total del ciclo.

Fortalezas identificadas

Como fortalezas se identifican la existencia de un marco normativo claro y consolidado que respalda las actuaciones de la entidad (Ley 909 de 2004, Decreto 1083 de 2015, Ley 1755 de 2015), la disponibilidad de información estructurada del aspirante en la plataforma SIMO y en archivos CSV, la experiencia acumulada del equipo de funcionarios en la elaboración de respuestas con soporte jurídico, la existencia de plantillas institucionales para la estructura de las comunicaciones, y el respaldo institucional para la implementación de soluciones tecnológicas, evidenciado en la autorización formal para el desarrollo del proyecto (ver Anexo A).

Oportunidades de mejora

Como oportunidades de mejora se identifican la necesidad de centralizar y versionar el repositorio normativo para facilitar la consulta y evitar la dispersión de fuentes, la conveniencia de estandarizar la argumentación jurídica mediante plantillas que aseguren uniformidad en las citas y reduzcan la variabilidad entre funcionarios, la posibilidad de automatizar la extracción de la solicitud del aspirante a partir del texto de la petición para reducir el esfuerzo manual de lectura e interpretación, la oportunidad de incorporar mecanismos de trazabilidad que vinculen cada respuesta con los fragmentos normativos que la sustentan, y la pertinencia de implementar un sistema de evaluación automática de la calidad del borrador antes de la revisión jurídica para reducir las devoluciones.

Procesamiento estadístico de datos

Los datos cuantitativos obtenidos de la encuesta tipo Likert se sistematizaron mediante estadística descriptiva, calculando medias, desviaciones estándar y frecuencias por ítem y por dimensión. La consistencia interna del instrumento se evaluó mediante el coeficiente Alfa de Cronbach por dimensión y a nivel global, con un umbral de aceptación de 0,70 (Bernal, 2020). Los resultados se presentan gráficamente mediante un diagrama de radar que permite visualizar el perfil de medias por dimensión, complementado con gráficos de barras a nivel de ítem para identificar los aspectos mejor y peor valorados, y una distribución global de las respuestas en la escala Likert.

Los datos cualitativos derivados del cuestionario de preguntas abiertas se analizaron mediante análisis categorial, codificando las respuestas en categorías emergentes (limitaciones, fortalezas, oportunidades, barreras) y triangulándolas con los resultados cuantitativos para identificar patrones convergentes y divergentes. Las métricas técnicas del agente, tales como tiempos de procesamiento, número de evidencias recuperadas y calificaciones del evaluador automático, se extrajeron de los registros de sesión generados por el sistema y se utilizaron para contextualizar los hallazgos perceptuales.

Análisis de los resultados

El análisis de resultados se organiza en cuatro apartados. El primero presenta el perfil de los participantes y la confiabilidad del instrumento. El segundo presenta los resultados cuantitativos de la encuesta Likert por dimensión y por ítem, incluyendo el perfil global de medias. El tercero presenta el análisis cualitativo del cuestionario de preguntas abiertas, organizado por bloques temáticos. El cuarto realiza la triangulación entre ambas fuentes y discute los hallazgos principales en relación con los objetivos del proyecto.

Perfil de los participantes y confiabilidad del instrumento

La encuesta Likert fue diligenciada por cinco funcionarios entre el 27 y el 30 de abril de 2026, con tiempos individuales de respuesta entre 4 y 39 minutos. Los participantes se desempeñan como contratistas y profesionales del Grupo de Gerencia de Convocatoria 12, vinculado al Proceso de Selección DIAN 2676. El cuestionario de preguntas abiertas fue diligenciado por dos funcionarios, cuyas respuestas se contrastaron con los resultados cuantitativos de la encuesta.

La consistencia interna del instrumento se evaluó mediante el coeficiente Alfa de Cronbach. La Tabla 4 presenta los valores obtenidos por dimensión y a nivel global. El Alfa global del instrumento (25 ítems) fue de 0,86, valor considerado bueno (Bernal, 2020). No obstante, dado el tamaño muestral del piloto ($N = 5$), la inferencia psicométrica se restringe explícitamente al Alfa global, y los coeficientes por dimensión no se interpretan como medidas de confiabilidad: con $N = 5$ y dimensiones de tres a cuatro ítems, el Alfa es altamente inestable y puede arrojar valores cercanos a cero o negativos (como ocurre en D4 (-1,30), D5 (0,07) y D6 (0,27)), que no expresan baja consistencia interna sino varianza inter-ítem negativa propia de muestras muy pequeñas. Para dimensionar esta inestabilidad se estimó un intervalo de confianza por remuestreo (bootstrap, 5.000 réplicas sobre los participantes) para el Alfa global, que resultó muy amplio: IC 95 % [0,30; 0,93]. Coeficientes alternativos más robustos, como el omega de McDonald o la confiabilidad compuesta, no son estimables de forma fiable bajo este diseño, pues un modelo factorial con $N = 5$ y 25 ítems queda subidentificado. En consecuencia, el estudio se caracteriza como exploratorio y los hallazgos por dimensión se reportan únicamente a partir de las medias y desviaciones estándar, sin pretender generalizaciones inferenciales.

Dimensión	Ítems (k)	N	Alfa
D1. Apoyo en identificación normativa y estructuración	4	5	0,56
D2. Calidad y precisión de citas normativas	4	5	0,72
D3. Eficiencia y reducción de tiempos	3	5	0,94
D4. Calidad de redacción y estructura argumentativa	3	5	-1,30
D5. Comprensión e interpretación de la petición	3	5	0,07
D6. Experiencia de usuario (UX) e interfaz	4	5	0,27
D7. Confianza, revisión humana y disposición de adopción	4	5	0,83
Global	25	5	0,86

Tabla 4 Coeficiente Alfa de Cronbach por dimensión y global. Elaboración propia a partir de los datos del piloto.

Nota. Los valores de Alfa cercanos a cero o negativos en dimensiones con $N = 5$ y $k = 3-4$ ítems reflejan la inestabilidad del coeficiente ante muestras muy pequeñas y no deben

interpretarse como ausencia de consistencia conceptual del constructo. La interpretación psicométrica se restringe al Alfa global (0,86; IC 95 % bootstrap [0,30; 0,93]).

Resultados cuantitativos de la encuesta Likert

Perfil global y por dimensión

La media global del instrumento (todas las respuestas, $n = 125$) fue de 3,11 puntos, con una desviación estándar de 1,15. Esta media se ubica ligeramente por encima del punto neutro de la escala (3 = ni de acuerdo ni en desacuerdo), lo que indica que la valoración del agente por parte de los funcionarios es moderadamente favorable, con una notable variabilidad entre ítems y entre participantes.

La distribución global de las 125 respuestas Likert (cinco participantes por 25 ítems) se concentra en las categorías centrales y de acuerdo: 8,8 % totalmente de acuerdo, 33,6 % de acuerdo, 27,2 % ni de acuerdo ni en desacuerdo, 20,8 % en desacuerdo y 9,6 % totalmente en desacuerdo. La siguiente figura presenta esta distribución.

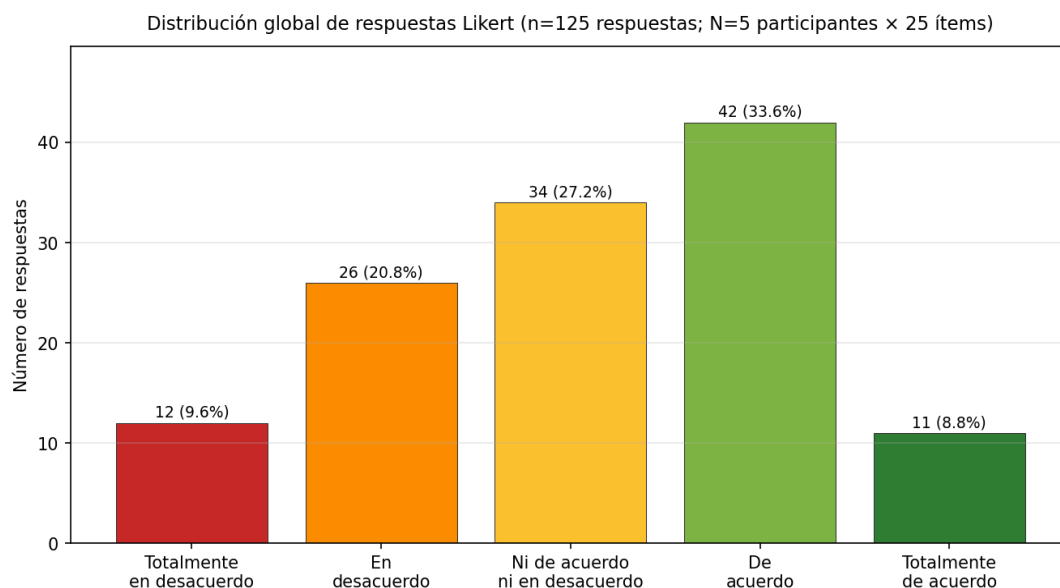


Figura 7 Distribución global de respuestas en la escala Likert. Elaboración propia a partir de los datos del piloto

Al agrupar los ítems por dimensión, las medias se ubican en un rango de 2,40 a 3,45, lo que se traduce en un perfil mixto: con dos dimensiones por encima de 3,30 (D2 y D7), tres dimensiones cercanas al punto neutro (D1, D4 y D6 entre 3,00 y 3,33), y dos dimensiones por

debajo del punto neutro (D3 y D5 con 2,40 y 2,93). A continuación, se presenta el perfil de medias por dimensión en formato radar y lo complementa en formato de barras horizontales.

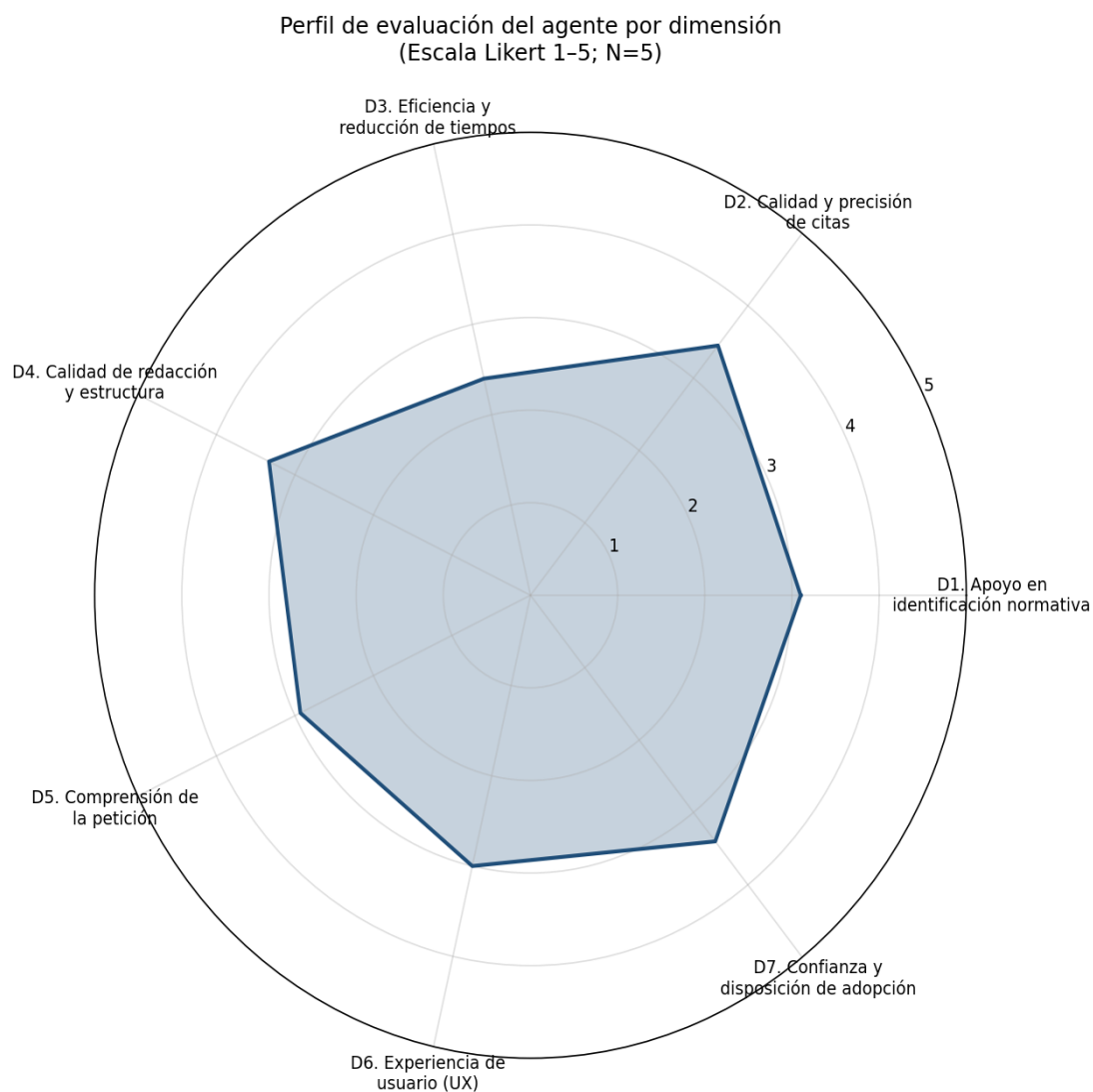


Figura 8 Perfil de medias por dimensión (diagrama radar, escala 1–5). Elaboración propia a partir de los datos del piloto.

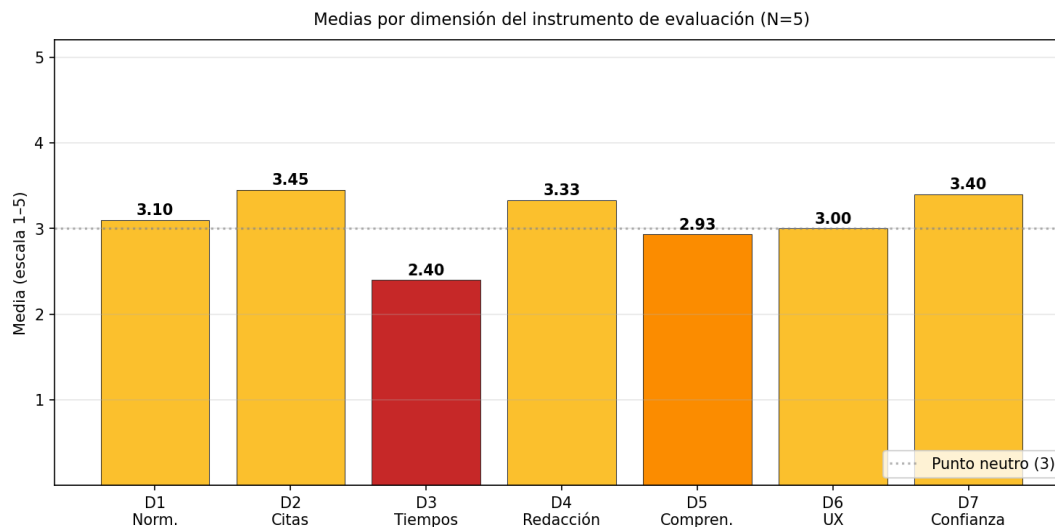


Figura 9 Medias por dimensión del instrumento de evaluación. Elaboración propia a partir de los datos del piloto.

La tabla sistematiza las medias y desviaciones estándar por dimensión, ordenadas de mayor a menor.

Dimensión	Media	DE
D2. Calidad y precisión de citas normativas	3,45	0,76
D7. Confianza, revisión humana y disposición de adopción	3,40	1,19
D4. Calidad de redacción y estructura argumentativa	3,33	0,98
D1. Apoyo en identificación normativa y estructuración	3,10	0,97
D6. Experiencia de usuario (UX) e interfaz	3,00	1,34
D5. Comprensión e interpretación de la petición	2,93	0,80
D3. Eficiencia y reducción de tiempos	2,40	1,55

Tabla 5 Medias y desviaciones estándar por dimensión. Elaboración propia a partir de los datos del piloto.

Resultados a nivel de ítem: aspectos mejor y peor valorados

Al desagregar a nivel de ítem se identifican patrones más nítidos. Se presenta las medias de los 25 ítems en orden de aparición en el instrumento, codificadas por color según la valoración.

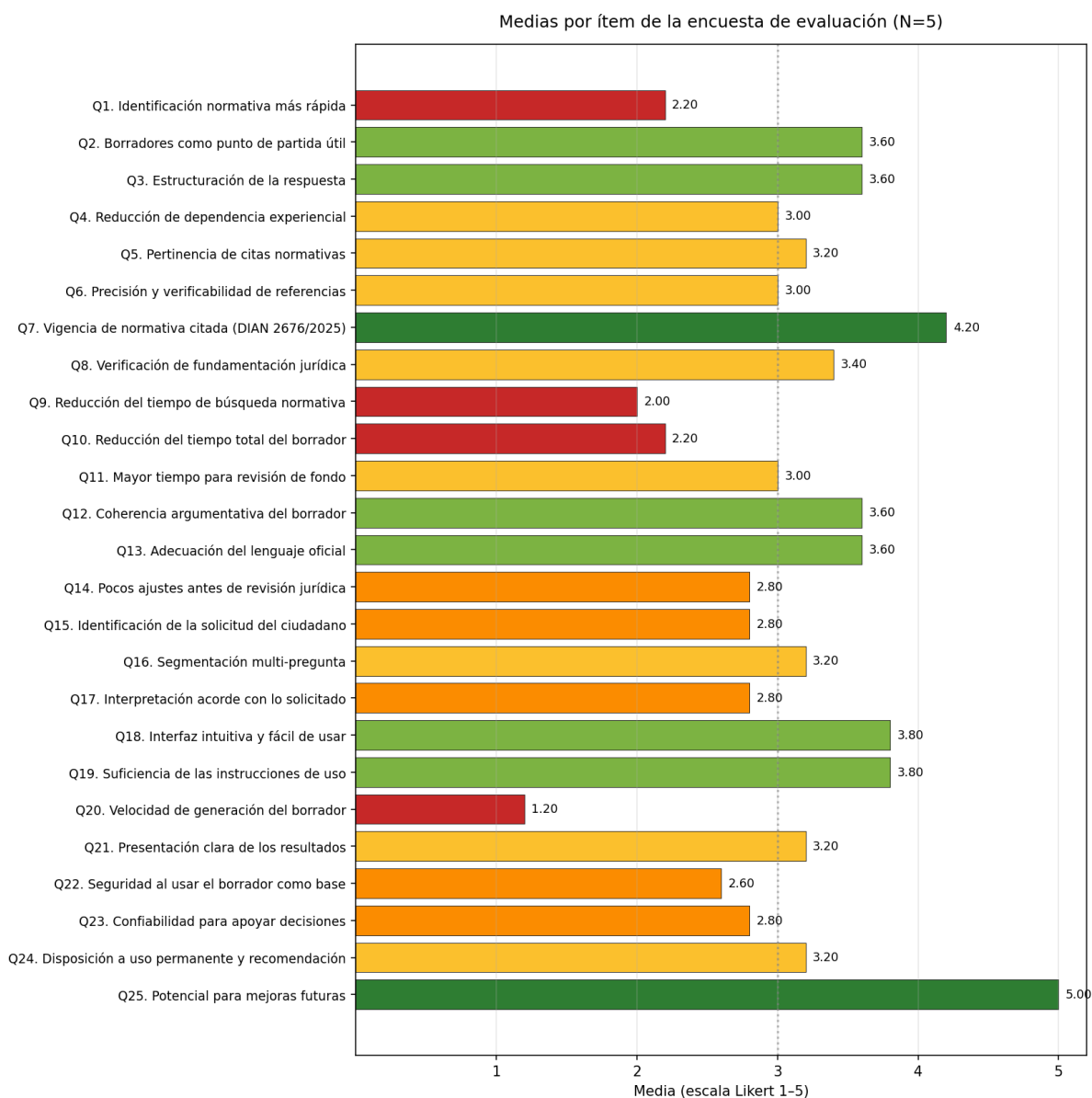


Figura 10 Medias por ítem de la encuesta de evaluación. Elaboración propia a partir de los datos del piloto.

Los cinco ítems con valoración más favorable son: el potencial del agente para mejorar la calidad de las respuestas e incorporar funcionalidades futuras (ítem Q25, cuyo enunciado literal es: "Considero que el agente tiene potencial para mejorar la calidad de las respuestas a peticiones e incorporar funcionalidades adicionales en el futuro"; $M = 5,00$; $DE = 0,00$); la vigencia y aplicabilidad de la normativa citada al Proceso de Selección DIAN 2676 ($M = 4,20$; $DE = 0,45$); la suficiencia de las instrucciones de capacitación para operar el agente ($M = 3,80$; $DE = 0,84$); la intuitividad de la interfaz ($M = 3,80$; $DE = 1,10$); y la utilidad del borrador como punto de partida ($M = 3,60$; $DE = 0,55$), empatado con la facilitación de la estructura de la

respuesta ($M = 3,60$; $DE = 0,55$), la coherencia argumentativa ($M = 3,60$; $DE = 0,89$) y la adecuación del lenguaje oficial ($M = 3,60$; $DE = 1,14$).

Los cinco ítems con valoración más desfavorable son: la velocidad de generación del borrador ($M = 1,20$; $DE = 0,45$); la reducción del tiempo de búsqueda normativa ($M = 2,00$; $DE = 1,22$); la rapidez en la identificación normativa frente al proceso manual ($M = 2,20$; $DE = 1,30$); el tiempo total para elaborar el borrador ($M = 2,20$; $DE = 1,64$); y la sensación de seguridad para usar el borrador como base sin revisión profunda ($M = 2,60$; $DE = 0,55$). Es relevante señalar que los cuatro primeros ítems pertenecen a las dimensiones D3 (eficiencia) y D1 (rapidez de identificación), evidenciando que la principal debilidad percibida está concentrada en el desempeño temporal del prototipo, no en la calidad del contenido generado.

Conviene interpretar con cautela la unanimidad perfecta del ítem Q25 ($M = 5,00$; $DE = 0,00$). Su enunciado es prospectivo (indaga por el potencial futuro del agente, no por su desempeño actual) y, por tanto, es susceptible a sesgo de aquiescencia y de deseabilidad social, especialmente en un piloto interno donde los participantes conocen el carácter institucional del desarrollo. Esta lectura se refuerza por el contraste con los ítems de desempeño efectivo, que recibieron valoraciones bajas (eficiencia D3, $M = 2,40$; velocidad de generación, $M = 1,20$). En consecuencia, este resultado no debe leerse como evidencia empírica del rendimiento del prototipo, sino como una actitud favorable de los funcionarios hacia la línea de trabajo y su continuidad. Para mitigar este sesgo en aplicaciones futuras del instrumento se recomienda equilibrar los ítems prospectivos con ítems de redacción invertida y separar explícitamente la medición del desempeño actual de la valoración del potencial.

Variabilidad entre participantes

Las medias individuales por participante oscilaron entre 2,48 y 3,68. La participante con la valoración global más baja ($M = 2,48$) reportó como cargo Ingeniera de Convocatoria, mientras que las valoraciones más altas correspondieron a contratistas que utilizan el agente como apoyo a la elaboración de borradores. Esta heterogeneidad se interpreta como evidencia del ajuste diferenciado del prototipo a distintos perfiles funcionales y a distintos tipos de petición, hallazgo que es coherente con la literatura sobre adopción de IA en organizaciones públicas (Neumann et al., 2022).

Análisis cualitativo del cuestionario de preguntas abiertas

Los dos cuestionarios cualitativos diligenciados aportan profundidad interpretativa a los hallazgos cuantitativos. El análisis se organiza por los bloques temáticos del instrumento.

Experiencia de uso del agente durante el piloto

Los dos participantes coinciden en valorar el agente como una herramienta con potencial, pero con limitaciones operativas en su versión actual. El primer participante describe la herramienta como útil pero lenta e imprecisa, valorando positivamente que presente un borrador con estructura lógica como punto de partida; sobre la calidad de las citas normativas señala que, aunque se extraen del acuerdo y el anexo, no fueron del todo precisas y estuvieron incompletas frente a la instrucción dada. La segunda participante reporta una experiencia donde la herramienta tiene potencial, pero las respuestas son inconclusas y los tiempos de espera son largos, incluso siendo específica con la pregunta y proporcionando el artículo base; valora la pertinencia de las citas normativas, pero no la coherencia entre la cita y la respuesta generada.

Manejo de peticiones multi-pregunta

Sobre la segmentación de peticiones multi-pregunta, el primer participante percibe que la herramienta identificó bien las inquietudes, pero no fue del todo precisa al argumentar las respuestas. La segunda participante reporta un caso en el que el agente no generó respuesta tras 20 minutos, lo que ilustra el problema de tiempos de inferencia en consultas complejas que evidencia también el ítem cuantitativo Q20 (velocidad).

Impacto en la eficiencia y la calidad del proceso

Ambos participantes reportan que el agente no redujo el tiempo de elaboración del borrador y que en algunos casos lo extendió. El primer participante señala que la imprecisión normativa genera desconfianza para usar el borrador sin una revisión previa suficiente, lo que limita la practicidad para reducir tiempos. La segunda participante es más enfática y reporta que los tiempos empeoran y que las respuestas dejan vacíos. Sobre la calidad final, el primer participante reconoce que la herramienta aporta una referencia de estructura útil; la segunda señala que el borrador puede ser un punto de partida pero las respuestas no son precisas frente a las preguntas.

Barreras, limitaciones y oportunidades de mejora

Las limitaciones convergentes identificadas son la imprecisión en la normativa citada y la lentitud para emitir el borrador. Las funcionalidades deseadas para versiones futuras incluyen

mayor velocidad y precisión en la generación, así como mayor capacidad para encontrar varias fuentes normativas que hagan más efectivo el trabajo. Las barreras de adopción identificadas son explícitamente los tiempos de respuesta y las respuestas inconclusas o no precisas frente a la pregunta del ciudadano.

Valoración general y recomendaciones

La valoración general en escala de 1 a 10 fue de 6 y de 4. El primer participante condiciona la recomendación del agente a otros equipos de la CNSC a la realización de ajustes en la precisión de la normativa citada y la velocidad de generación. La segunda participante no recomendaría su uso por ahora, dado que se trata de una versión beta a la que le faltan cosas por mejorar e implementar. Como observación final, el primer participante reconoce el desarrollo como un avance con potencial considerable frente al alto flujo de peticiones del equipo.

Triangulación y discusión de hallazgos

La triangulación entre los resultados cuantitativos y cualitativos revela tres patrones convergentes. El primer patrón es la consistencia entre la baja valoración Likert de los ítems de eficiencia (D3, M = 2,40; Q20 sobre velocidad, M = 1,20) y la queja recurrente en las respuestas abiertas sobre la lentitud del agente. El segundo patrón es la convergencia entre la valoración intermedia de la calidad de citas (D2, M = 3,45) y la observación cualitativa de que las citas normativas son extraídas de las fuentes correctas pero presentan incompletitudes y desajustes argumentativos en algunos casos, resultado consistente con los hallazgos de Pipitone y Houir Alami (2024) sobre la criticidad del componente de recuperación en pipelines RAG aplicados al dominio jurídico. El tercer patrón es la consistencia entre la valoración favorable de potencial futuro (Q25, M = 5,00; D7 disposición, M = 3,40) y la observación cualitativa de que la herramienta tiene un enorme potencial frente al alto flujo de peticiones del equipo, condicionado a mejoras en velocidad y precisión.

Estos hallazgos se enmarcan en la literatura sobre adopción de IA en el sector público. La preferencia por mantener al humano en el ciclo, observada explícitamente en el ítem Q22 (M = 2,60, neutro-bajo) y en las respuestas abiertas que insisten en la necesidad de revisión previa antes del envío, es consistente con la evidencia experimental de Sele y Chugunova (2024) sobre la importancia de la supervisión humana para la aceptación de decisiones algorítmicas. La convergencia de la crítica sobre tiempos con la fortaleza percibida en estructura y vigencia normativa sugiere que el cuello de botella del prototipo no está en su lógica funcional sino en la

infraestructura de inferencia: el modelo de lenguaje qwen2.5:3b ejecutado exclusivamente sobre CPU en una instancia ARM Ampere A1 introduce latencias que el equipo de funcionarios percibe como inaceptables, conclusión coherente con la evidencia sobre ineficiencias en pipelines RAG bajo restricciones de cómputo, donde la inferencia exclusiva en CPU puede duplicar las latencias respecto a configuraciones con aceleración por GPU.

Es necesario reconocer las limitaciones del estudio. El tamaño muestral del piloto (N = 5 para la encuesta, n = 2 para el cuestionario abierto) restringe la generalización de los hallazgos y la estabilidad estadística del Alfa de Cronbach por dimensión, aunque el Alfa global (0,86) avala la consistencia conceptual del instrumento. La duración del piloto (18 días de uso activo del agente, del 13 al 30 de abril de 2026, con aplicación de instrumentos de evaluación los días 27 y 28 de abril) y la concentración en un único proceso de selección (DIAN 2676) limitan también la inferencia sobre desempeño en otros contextos. No obstante, los hallazgos cualitativos son ricos y convergen con la evidencia cuantitativa, lo que aporta validez interpretativa al diagnóstico.

En cuanto a la situación actual del proceso, los datos del piloto confirman que las debilidades operativas identificadas en el diagnóstico organizacional, con un tiempo de elaboración del borrador de 3,4 días hábiles que representa poco más de la mitad del ciclo, no se resuelven con el prototipo en su configuración actual. Las fortalezas que el agente potencia son la consistencia argumentativa, en estructura y lenguaje oficial, y la trazabilidad normativa, en vigencia y aplicabilidad de las citas, mientras que las oportunidades de mejora se concentran en el desempeño temporal del modelo y en la profundidad argumentativa entre la cita y el razonamiento.

Plan de Intervención

El plan de intervención consiste en el desarrollo e implementación de un agente de inteligencia artificial orientado a apoyar la elaboración de respuestas a peticiones ciudadanas en la CNSC. La solución se estructura como un prototipo funcional (producto mínimo viable) que integra los componentes descritos a continuación.

Arquitectura de la solución

El agente se basa en la arquitectura de Generación Aumentada por Recuperación (RAG), que combina un componente de búsqueda semántica sobre un repositorio normativo

versionado con un componente de generación de texto que redacta borradores a partir de los fragmentos normativos recuperados. La arquitectura se organiza en los siguientes módulos:

El primer módulo corresponde al repositorio normativo controlado. Se consolidó un corpus que incluye los acuerdos de convocatoria, anexos y normatividad aplicable a los procesos de selección de la CNSC. Cada documento fue procesado y segmentado en fragmentos mediante un algoritmo de segmentación jurídica que identifica unidades por artículo, numeral, párrafo, título, capítulo y sección, con un tamaño máximo de 800 tokens y un solapamiento de 150 tokens entre fragmentos. Los fragmentos fueron indexados con metadatos de fuente, fecha y numeral. El corpus se almacena en un índice vectorial construido con FAISS y un índice léxico complementario, lo que permite realizar búsquedas híbridas, semánticas y léxicas, sobre la normativa (Johnson et al., 2019).

El segundo módulo es el pipeline de recuperación normativa. El sistema implementa una búsqueda híbrida que combina recuperación semántica mediante embeddings generados con el modelo sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 y recuperación léxica con fusión mediante Reciprocal Rank Fusion (RRF). Los resultados son reordenados por un cross-encoder (mmarco-mMiniLMv2-L12-H384-v1) para mejorar la precisión. Adicionalmente, el sistema incluye detección automática de complejidad de la consulta, generación de múltiples queries por consulta y un límite máximo de dos evidencias por documento fuente para evitar redundancia.

La búsqueda híbrida responde a una necesidad identificada durante el desarrollo del prototipo: la búsqueda exclusivamente semántica puede omitir coincidencias exactas de numerales, artículos y nombres de acuerdos que son frecuentes en las consultas de los funcionarios, mientras que la búsqueda exclusivamente léxica no captura relaciones de significado entre términos jurídicos equivalentes. La combinación de ambos enfoques mediante RRF con un parámetro $k=60$, seguida del reranking con cross-encoder, permite obtener fragmentos que son tanto semánticamente pertinentes como textualmente precisos en la referencia normativa. Los parámetros de configuración establecen $k=6$ resultados finales, con un pool de candidatos de 300 para la búsqueda semántica y 400 para la léxica, lo que asegura una exploración amplia antes del filtrado final.

El tercer módulo corresponde a la segmentación de peticiones y generación de borradores. El sistema incluye un segmentador de peticiones multi-pregunta que detecta subsolicitudes mediante conectores aditivos, enumeraciones, signos de interrogación y frases de solicitud explícitas (como "solicito", "requiero" o "deseo saber"), y procesa cada segmento de forma independiente para asegurar respuesta de fondo a cada solicitud del ciudadano. El cruce

con los datos del aspirante se realiza, en el prototipo, sobre un archivo estructurado (CSV/XLSX) exportado del SIMO, mediante búsqueda por número de cédula y bajo el principio de minimización de datos, de modo que la información personal permanece local y no se transmite a servicios externos; la integración en línea con el SIMO se plantea como trabajo futuro. Utilizando los fragmentos recuperados como contexto, un modelo de lenguaje (qwen2.5:3b ejecutado localmente mediante Ollama) genera el borrador de respuesta, incorporando las citas normativas de forma explícita. El sistema aplica plantillas institucionales que definen la estructura del documento de salida y puede utilizar modelos de respuesta reutilizables como referencia para el LLM.

El cuarto módulo es el evaluador automático de calidad. Antes de presentar el borrador al funcionario, el sistema ejecuta un checklist de ocho verificaciones objetivas que evalúan: la presencia de citas normativas con referencia a evidencia, la especificación de artículos y numerales, la transcripción de la petición del ciudadano entre comillas, la existencia de secciones estructuradas (antecedentes, consideraciones, fundamento y decisión), la longitud suficiente del borrador, la ausencia de campos pendientes, la diversidad de fuentes normativas consultadas y la ausencia de conectores redundantes repetidos. El resultado se presenta como una calificación (A, B, C o D) con sugerencias de mejora.

El quinto módulo es la interfaz de usuario. El agente se despliega a través de una aplicación web desarrollada en Streamlit, que permite al funcionario ingresar el texto de la petición, visualizar los fragmentos normativos recuperados con sus niveles de confianza, revisar el borrador generado junto con la evaluación de calidad, y exportar la respuesta en formato documental. La interfaz incluye un panel de evidencias que muestra las fuentes consultadas. Adicionalmente, el sistema registra logs estructurados de cada sesión (query, evidencias, evaluación, tiempos) para el análisis posterior del piloto.

Stack tecnológico

La solución se construye sobre tecnologías abiertas y de bajo costo, organizadas en una arquitectura modular que facilita el mantenimiento y la escalabilidad. La Tabla 6 presenta los componentes principales del stack tecnológico empleado en el prototipo.

Componente	Tecnología	Función
Lenguaje de desarrollo	Python 3.10+	Desarrollo de todos los módulos del agente

Indexación vectorial	FAISS	Búsqueda por similitud semántica sobre el corpus normativo
Embeddings	sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2	Generación de representaciones vectoriales de texto en español
Reranking	cross-encoder/mmarco-mMiniLMv2-L12-H384-v1	Reordenamiento de resultados por relevancia
Modelo de lenguaje	qwen2.5:3b (vía Ollama)	Generación de borradores de respuesta
Interfaz web	Streamlit	Interfaz para el funcionario
Infraestructura	Oracle Cloud Infrastructure (VM.Standard.A2.Flex)	Despliegue como servicio persistente (systemd)

Tabla 6 Stack tecnológico del agente. *Elaboración propia a partir de la configuración del prototipo funcional.*

La selección de estas tecnologías responde a criterios de costo, rendimiento y facilidad de mantenimiento por equipos internos. El modelo de embeddings multilingüe permite procesar texto en español sin necesidad de traducción previa. El cross-encoder aporta precisión en el reordenamiento de fragmentos normativos, mientras que el modelo de lenguaje qwen2.5:3b, ejecutado localmente mediante Ollama, permite generar texto sin depender de servicios externos, lo cual es relevante para la seguridad y confidencialidad de la información institucional.

Justificación de la selección del modelo de lenguaje y la arquitectura de inferencia

La selección del modelo de lenguaje qwen2.5:3b y de una arquitectura de inferencia local respondió a un conjunto explícito de criterios, priorizados en el siguiente orden. El primero es la protección de datos personales y la reserva legal: las peticiones contienen datos de aspirantes (identificación, hoja de vida, condición de discapacidad, entre otros) sujetos a la Ley 1581 de 2012, por lo que se descartó el envío de información gubernamental a interfaces de programación (API) comerciales externas y se optó por una ejecución local que garantiza la soberanía del dato. El segundo es el costo: el desarrollo se realizó sobre la capa gratuita y de bajo costo de Oracle Cloud Infrastructure, en coherencia con la ausencia de infraestructura especializada de cómputo en la entidad al momento del proyecto. El tercero es el despliegue autónomo y mantenible por equipos internos, sin dependencia de proveedores externos. El cuarto es la viabilidad técnica en el hardware disponible: un modelo de 3.000 millones de

parámetros es ejecutable sobre CPU en una instancia ARM Ampere A1, mientras que modelos mayores resultan inviables sin aceleración por GPU.

A la luz de estos criterios se evaluaron y descartaron varias alternativas. Los modelos comerciales accedidos por API (por ejemplo, los de OpenAI o Anthropic) ofrecen mayor capacidad de razonamiento, pero implican el envío de datos sensibles a terceros y un costo recurrente por token, incompatibles con la reserva legal y con la restricción presupuestal; aun con cláusulas de no entrenamiento, el flujo de datos fuera de la entidad se consideró un riesgo no asumible en esta fase. Los modelos abiertos de mayor tamaño (7B–14B), incluso cuantizados, y los servidores de inferencia optimizada como vLLM o TGI mejoran la latencia y la calidad, pero requieren GPU para operar con tiempos aceptables, infraestructura no disponible en la capa utilizada. En consecuencia, qwen2.5:3b sobre CPU se adoptó como una decisión de viabilidad para el prototipo (producto mínimo viable), asumiendo conscientemente la limitación de desempeño temporal que el piloto efectivamente evidenció. La siguiente tabla sintetiza la comparación.

Alternativa	Latencia esperada	Costo	Privacidad / soberanía del dato	Viabilidad en el proyecto
qwen2.5:3b en CPU (ARM Ampere A1, OCI)	Alta (lenta; cuello de botella del piloto)	Muy bajo (capa gratuita)	Alta (ejecución local, sin envío a terceros)	Adoptada (PMV)
Modelo abierto 7B-14B cuantizado en GPU (AWS, Azure u OCI)	Baja (segundos por borrador)	Medio (GPU A10 ≈ 1,0–3,2 US\$/hora; ≈ 180–560 US\$/mes en jornada laboral)	Recomendada para producción	Recomendada para producción
Servidor de inferencia optimizada (vLLM/TGI) en GPU	Baja	Medio (requiere GPU)	Alta (local)	Recomendada (con lo anterior)
API comercial (OpenAI/Anthropic)	Muy baja	Variable por token	Baja (envío a terceros; riesgo)	Descartada

			frente a Ley 1581)	
--	--	--	-----------------------	--

Tabla 7 Comparación de alternativas de modelo e inferencia según latencia, costo y privacidad. Elaboración propia.

En términos de continuidad, llevar la inferencia a producción exige una instancia con aceleración por GPU en un servicio de nube administrado, dado que la entidad no dispone de infraestructura ni de servicios de nube propios. El costo de cómputo de una GPU de clase NVIDIA A10 oscila, según el proveedor y la región, entre aproximadamente 1,0 y 3,2 dólares por hora en la modalidad bajo demanda; por ejemplo, AWS g5.xlarge con A10G alrededor de 1,0 dólar por hora, Oracle Cloud Infrastructure VM.GPU.A10 alrededor de 2,0 dólares por hora y Microsoft Azure de la serie NVads A10 v5 entre 0,9 y 3,2 dólares por hora según el tamaño. En un esquema de uso acotado a la jornada laboral, cercano a 176 horas al mes, el costo mensual se ubica aproximadamente entre 180 y 560 dólares, mientras que en operación continua, de 730 horas al mes, se ubica entre 730 y 2.300 dólares (Amazon Web Services, 2026; Microsoft, 2026; Oracle, 2026). Esta cifra es el insumo económico central para que la CNSC decida sobre la continuidad descrita en la primera recomendación, y debe contrastarse con el costo de oportunidad de los 3,4 días hábiles promedio que hoy demanda la elaboración manual de cada borrador.

Se presenta la arquitectura general del agente, mostrando el flujo de datos desde la recepción de la petición hasta la generación del borrador con citas trazables.

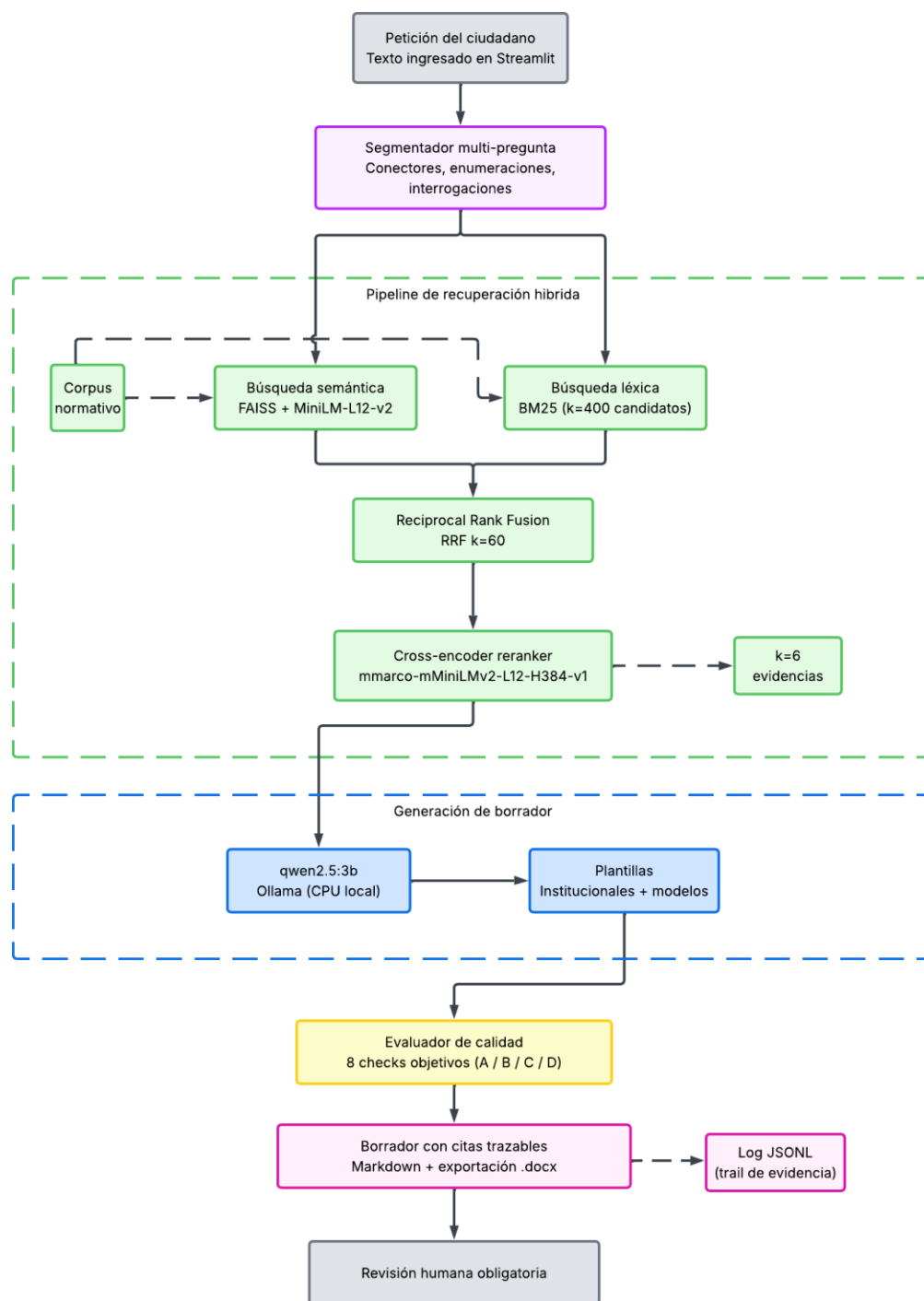


Figura 11 Arquitectura del agente RAG para la elaboración asistida de respuestas a peticiones. Elaboración propia.

Protocolo del piloto

La validación del agente se desarrolló mediante un piloto controlado con el equipo del Grupo de Gerencia de Convocatoria 12 vinculado al Proceso de Selección DIAN 2676. El protocolo se ejecutó en tres fases.

En la primera fase se proporcionó al equipo acceso al agente desplegado y se realizó una sesión de capacitación sobre su uso, alcances y limitaciones. Se enfatizó que el agente opera como herramienta de apoyo y que la revisión humana del borrador es obligatoria antes del envío de cualquier respuesta oficial.

En la segunda fase los funcionarios utilizaron el agente para elaborar borradores de respuesta a peticiones del proceso. Durante este periodo se registraron sesiones del sistema con sus respectivas consultas, evidencias y evaluaciones automáticas.

En la tercera fase se aplicaron los instrumentos de medición. La encuesta tipo Likert fue diligenciada por cinco funcionarios y el cuestionario de preguntas abiertas fue diligenciado por dos funcionarios. Los resultados se analizaron conforme a lo descrito en las secciones de procesamiento estadístico y análisis de resultados.

Entregables del plan de intervención

Los entregables asociados al plan de intervención se organizan en dos categorías: entregables técnicos y entregables de validación. En la categoría técnica se incluyen el agente funcional desplegado y accesible para el equipo de la CNSC, el repositorio normativo versionado con metadatos de fuente y vigencia, el índice vectorial FAISS con los embeddings del corpus normativo, las plantillas institucionales para la generación de respuestas con campos parametrizables, y los modelos de respuesta reutilizables que sirven como referencia para el LLM en la generación de borradores.

En la categoría de validación se incluyen la encuesta estructurada tipo Likert con las siete dimensiones de evaluación, el cuestionario de preguntas abiertas para la indagación cualitativa y los resultados del piloto efectivamente realizado, presentados en este informe.

Plan de implementación (propuesta para producción)

Esta sección presenta una propuesta de carácter prospectivo: la ruta para llevar el prototipo, validado durante el piloto, a un entorno de producción en la CNSC. No describe actividades ya ejecutadas, sino las fases, los responsables y los requisitos necesarios para su puesta en operación. Su inclusión es coherente con el carácter de Trabajo Dirigido del proyecto, orientado a la transferencia de la solución.

Fases y cronograma

Se proponen cuatro fases. (1) Estabilización técnica (mes 1-2): migración de la inferencia a una instancia con GPU o sustitución del modelo por una variante optimizada, hasta alcanzar tiempos de generación aceptables (objetivo de referencia: por debajo de un minuto por borrador). (2) Endurecimiento y seguridad (mes 2-3): controles de acceso, registro de auditoría, respaldo del corpus y de los índices, y verificación del tratamiento de datos personales conforme a la Ley 1581 de 2012. (3) Despliegue controlado y capacitación (mes 3-4): puesta en producción para el Grupo de Gerencia de Convocatoria, con capacitación y guía de uso. (4) Escalamiento (mes 4-6): extensión a otros grupos y procesos de selección, con seguimiento de indicadores.

Roles y responsabilidades

La operación sostenible requiere asignar formalmente los siguientes roles dentro de la entidad.

Rol	Responsabilidad principal
Patrocinador institucional (CNSC)	Decisión de continuidad, asignación de presupuesto y definición de prioridades.
Líder técnico / administrador del agente	Despliegue, operación, monitoreo y actualización del agente y de los índices.
Equipo jurídico	Validación de fondo de los borradores y definición de criterios de cita.
Gestión documental / TI	Integración con SIMO, respaldos, seguridad y control de versiones.
Funcionarios elaboradores	Uso del agente en el flujo de trabajo y retroalimentación para la mejora.

Tabla 8 Roles y responsabilidades para la operación del agente. *Elaboración propia.*

Requisitos y gestión del cambio

Los requisitos mínimos comprenden infraestructura de cómputo con aceleración por GPU, el acceso controlado al corpus normativo y a los datos del aspirante, y un procedimiento formal de actualización del repositorio normativo. Dado que la CNSC no dispone actualmente de infraestructura ni de servicios de nube propios, para producción se recomienda adoptar un servicio de nube administrado, como Amazon Web Services (AWS) o Microsoft Azure, que ofrecen instancias con GPU bajo demanda y servicios gestionados de seguridad y respaldo. La

gestión del cambio contempla la comunicación del alcance del agente como herramienta de apoyo con revisión humana obligatoria, la capacitación continua y un canal de retroalimentación para la mejora iterativa. El éxito de la implementación se medirá con los indicadores propuestos en el modelo de operación de la sección de recomendaciones.

Punto de partida: lo ejecutado en el piloto

El piloto efectivamente ejecutado (13 al 30 de abril de 2026) constituye la primera iteración de este plan: se desplegó el agente como servicio web persistente (systemd) en Oracle Cloud Infrastructure, se capacitó al equipo, se registraron las sesiones de uso (consulta, evidencias, evaluación automática y tiempos) y se aplicaron los instrumentos de validación, cuyos resultados se documentan en la sección de análisis de resultados. De este modo, el tercer objetivo específico (desarrollo y operación del agente) y el quinto (validación) quedan documentados con evidencia empírica del despliegue y del uso real por parte de los funcionarios.

Conclusiones y Recomendaciones

Conclusiones

A partir del trabajo realizado, se presentan las siguientes conclusiones, alineadas con los objetivos específicos del proyecto y con los hallazgos del piloto.

En relación con el primer objetivo, se logró consolidar un repositorio normativo controlado que integra los acuerdos de convocatoria, anexos y normatividad aplicable a los procesos de selección de la CNSC. Este repositorio fue segmentado mediante un algoritmo de segmentación jurídica, indexado con metadatos de fuente y numeral, y almacenado en un índice vectorial FAISS complementado con un índice léxico, lo que permite la consulta semántica e híbrida reproducible sobre la normativa vigente. La pertinencia y vigencia de las citas normativas para el Proceso de Selección DIAN 2676 fue el aspecto mejor valorado del agente en la encuesta ($M = 4,20$).

En relación con el segundo objetivo, se implementó un componente RAG con búsqueda híbrida (semántica y léxica), reranking mediante cross-encoder y controles de diversidad de fuentes. La calidad y precisión de las citas normativas obtuvo una valoración global moderadamente favorable ($M = 3,45$ a nivel de dimensión), con observaciones cualitativas que resaltan la pertinencia de las fuentes recuperadas pero señalan incompletitudes en la articulación entre cita y argumentación, hallazgo coherente con el problema reportado en la

literatura sobre la criticidad del componente de recuperación en pipelines RAG aplicados al dominio jurídico (Pipitone y Hour Alami, 2024).

En relación con el tercer objetivo, se desarrolló un agente que, a partir del texto de la petición, identifica la solicitud del ciudadano mediante un segmentador de peticiones multi-pregunta y genera un borrador de respuesta apoyado en la evidencia normativa recuperada. La utilidad del borrador como punto de partida fue valorada favorablemente ($M = 3,60$), al igual que la coherencia argumentativa ($M = 3,60$) y la adecuación del lenguaje oficial ($M = 3,60$). El evaluador automático de calidad opera correctamente y orienta al funcionario sobre los aspectos del borrador que requieren revisión adicional.

En relación con el cuarto objetivo, se estandarizó el formato de salida mediante plantillas institucionales que integran el borrador argumentado, las citas verificables y las evidencias normativas, con campos parametrizables y exportación a formato documental. La interfaz fue valorada como intuitiva y fácil de utilizar ($M = 3,80$), y las instrucciones de capacitación fueron consideradas suficientes ($M = 3,80$).

En relación con el quinto objetivo, se ejecutó el piloto controlado y se aplicaron los instrumentos diseñados (encuesta Likert con $N = 5$ y cuestionario de preguntas abiertas con $n = 2$). Los resultados muestran un perfil de aceptación mixto: alta valoración de la pertinencia y vigencia normativa (D2, D7) y de la utilidad del prototipo como punto de partida, pero baja valoración del desempeño temporal (D3, $M = 2,40$; ítem específico de velocidad $M = 1,20$). Esta debilidad se atribuye a la inferencia exclusivamente sobre CPU en una instancia ARM Ampere A1, sin aceleración por GPU, condición de infraestructura que constituye un hallazgo del proyecto y no una falla de diseño del agente. Es importante señalar que el potencial percibido del agente para mejoras futuras alcanzó la máxima valoración posible ($M = 5,00$; $DE = 0,00$), lo que indica un consenso pleno entre los participantes sobre el valor del enfoque.

Como conclusión transversal, el proyecto materializa un caso documentado de aplicación de IA generativa en una entidad pública colombiana, que articula desarrollo técnico, intervención organizacional y validación empírica. Los hallazgos confirman tanto la viabilidad técnica del enfoque RAG para apoyar la elaboración de respuestas a peticiones, como la centralidad del enfoque de humano en el ciclo: los funcionarios mantienen una actitud cautelosa frente al uso del borrador sin revisión profesional, lo que es coherente con la evidencia de que la presencia de supervisión humana es un determinante crítico de la confianza y aceptación de la IA en la administración pública (Sele y Chugunova, 2024; Agbabiaka et al., 2025).

Recomendaciones

A partir de los hallazgos del proyecto y del piloto, se formulan las siguientes recomendaciones para la implementación, mejora y continuidad de la solución.

La primera recomendación es priorizar la mejora del desempeño temporal del agente como condición habilitante para su adopción. Esto puede lograrse mediante la migración de la inferencia a una instancia con aceleración por GPU en un servicio de nube administrado, como Amazon Web Services o Microsoft Azure, dado que la entidad no cuenta con infraestructura propia, junto con la sustitución del modelo de lenguaje por una variante optimizada para inferencia, por ejemplo modelos cuantizados o servidos con vLLM, o la combinación de ambas estrategias. La evidencia del piloto muestra que la velocidad de generación es el cuello de botella perceptual más crítico; sin una mejora sustancial en este aspecto, la adopción permanente resulta improbable.

La segunda recomendación es fortalecer la articulación entre la cita normativa recuperada y el razonamiento jurídico expresado en el borrador. Esto implica revisar las plantillas de instrucción para favorecer una redacción que cite el numeral concreto y desarrolle el argumento a partir del fragmento recuperado, así como ampliar el límite de tokens para permitir respuestas argumentativas más completas ante peticiones de mayor complejidad. Las observaciones cualitativas sobre imprecisión y vacíos argumentativos apuntan directamente a esta necesidad.

La tercera recomendación es establecer un procedimiento periódico de actualización del repositorio normativo, de modo que la incorporación de nuevos acuerdos, anexos o modificaciones se refleje oportunamente en el índice vectorial y léxico. Este procedimiento debería formalizarse como parte del flujo institucional de gestión documental.

La cuarta recomendación es preservar y reforzar el enfoque de humano en el ciclo, haciendo explícito en cada borrador su carácter de propuesta sujeta a revisión profesional obligatoria antes del envío. Esta recomendación es coherente con los hallazgos sobre confianza (D7, M = 3,40) y con la literatura internacional sobre aceptación de IA en el sector público (Sele y Chugunova, 2024).

La quinta recomendación es ampliar el piloto a un mayor número de funcionarios y a otros procesos de selección de la CNSC, una vez atendidas las mejoras técnicas anteriores. Un piloto extendido permitiría obtener evidencia con mayor potencia estadística y detectar diferencias de adopción según el tipo de petición y el perfil del funcionario.

La sexta recomendación es avanzar hacia la integración del agente con los sistemas institucionales de la CNSC, en particular con SIMO, con el fin de automatizar el cruce de datos del aspirante y reducir la dependencia de archivos de carga manual. Esta integración facilitaría también la trazabilidad de las respuestas y la generación de reportes agregados sobre el uso de la herramienta.

La séptima recomendación es documentar las lecciones aprendidas durante el piloto e incorporarlas en una guía institucional de uso del agente, que sirva como referencia para la capacitación de nuevos funcionarios y para la eventual transferencia de la solución a otras entidades del sector público que enfrenten desafíos similares en la gestión de peticiones ciudadanas.

Modelo de operación, transferencia y sostenibilidad

Como Trabajo Dirigido orientado a la transferencia, el proyecto contempla las condiciones para que la solución opere y se sostenga en la entidad más allá de la tesis, incorporando la dimensión emprendedora y de valor público.

Modelo de operación post-piloto

Una vez finalizada la tesis, la operación del agente debe asignarse formalmente dentro de la CNSC. Se propone que un líder técnico o el área de TI asuma la operación, el mantenimiento y la actualización del agente y del corpus normativo; que el área jurídica conserve la validación de fondo; y que el patrocinador institucional defina el presupuesto anual y las prioridades. Como métricas de gestión para esa operación se proponen el tiempo promedio de elaboración del borrador, el número de peticiones atendidas con apoyo del agente y la satisfacción del funcionario. Estas métricas no se midieron en el presente estudio, de alcance exploratorio, sino que se plantean como instrumento de seguimiento para una eventual fase de producción. El presupuesto anual de referencia debe cubrir la infraestructura de cómputo en la nube, el mantenimiento del corpus y la capacitación, conforme al modelo de costos presentado más adelante.

Escalabilidad y replicabilidad hacia otras entidades

La arquitectura es replicable hacia otras entidades del sector público colombiano que gestionan grandes volúmenes de peticiones sobre marcos normativos extensos, como el Departamento Administrativo de la Función Pública, las superintendencias o las alcaldías. La replicación exige sustituir el corpus normativo por el de la entidad destino, ajustar las plantillas

institucionales y validar con sus funcionarios. Las principales barreras de adopción son la disponibilidad de infraestructura de cómputo, la madurez de la gestión documental y la gestión del cambio organizacional; los principales habilitadores son el carácter abierto de la solución y su independencia de proveedores externos.

Propiedad intelectual y licenciamiento

El código del agente se publica en un repositorio público de GitHub. Para habilitar su uso, adaptación y transferencia por parte de la CNSC y de otras entidades, su licenciamiento se declara bajo la licencia AGPL-3.0-or-later (Aferro General Public License). Esta elección obedece a dos razones. La primera es de compatibilidad de licencias: la dependencia PyMuPDF, empleada para el procesamiento de los documentos PDF del corpus normativo, se distribuye bajo AGPL-3.0, una licencia de copyleft fuerte que se propaga a las obras derivadas. La segunda es de principio: la AGPL garantiza que cualquier uso, modificación o despliegue como servicio en red conserve el código abierto y disponible, en coherencia con la transparencia propia del sector público. El proyecto se publica como trabajo académico, sin comprometer a la CNSC ni a la Universidad Ean, y el repositorio incluye el archivo de licencia, una política de seguridad y la declaración de autoría.

Modelo de costos operativos anualizados

El siguiente modelo presenta los costos operativos anualizados de referencia para la operación del agente en producción.

Rubro	Costo anual estimado (referencial)
Infraestructura de cómputo en la nube con GPU (AWS, Azure u OCI; clase A10)	≈ US\$2.200 - US\$27.600 anuales según intensidad de uso (≈ US\$180–560/mes en jornada laboral; hasta ≈ US\$2.300/mes en operación continua)
Mantenimiento y actualización del corpus normativo	Tiempo de personal interno (sin costo de licencias)
Capacitación y soporte a usuarios	Tiempo de personal interno
Licenciamiento de software	US\$0 (componentes de código abierto)

Tabla 9 *Modelo de costos operativos anualizados (referencial). Elaboración propia a partir de Oracle (2026).*

Los valores son referenciales y se basan en los precios públicos de lista de Amazon Web Services, Microsoft Azure y Oracle Cloud Infrastructure (Amazon Web Services, 2026; Microsoft, 2026; Oracle, 2026); deben ajustarse a la intensidad de uso real, a la región y a los

acuerdos de la entidad. El componente dominante es la infraestructura de cómputo; los demás rubros se cubren con recurso humano interno, dado el carácter abierto de la solución.

Referencias

Agbabiaka, O., Ojo, A. y Connolly, N. (2025). Requirements for trustworthy AI enabled automated decision-making in the public sector: A systematic review. *Technological Forecasting and Social Change*, 215, 124076. <https://doi.org/10.1016/j.techfore.2025.124076>

AGN. (2001). Acuerdo 060 de 2001. Pautas para la administración de comunicaciones oficiales. Archivo General de la Nación.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=10551>

AGN. (2024). Acuerdo 001 de 2024. Archivo General de la Nación.
<https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=154286>

Amazon Web Services. (2026). Amazon EC2 G5 Instances: pricing.
<https://aws.amazon.com/ec2/instance-types/g5/>

Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P. y Horvitz, E. (2019). Guidelines for human–AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3290605.3300233>

Ariai, F., Hettiarachchi, H., Adel, H., Scherp, A., Ruas, T. y Gipp, B. (2024). Natural language processing for the legal domain: A survey. *arXiv:2410.21306*.
<https://arxiv.org/abs/2410.21306>

Arias, F. (2012). *El proyecto de investigación científica* (6.^a ed.). Editorial Episteme.

Bernal, C. A. (2020). *Metodología de la investigación: para administración, economía, humanidades y ciencias sociales* (4.^a ed.). Pearson Educación.

Chiavenato, I. y Sapiro, A. (2019). *Planeación estratégica: fundamentos y aplicación* (3.^a ed.). McGraw-Hill.

CNSC. (2021). Acuerdo 2073 de 2021. Estructura organizacional de la Comisión Nacional del Servicio Civil.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=171347>

CNSC. (2022, 7 de octubre). La CNSC cumple 18 años como la entidad responsable de la administración y vigilancia de la carrera administrativa en Colombia. Comisión Nacional del Servicio Civil. <https://www.cnsc.gov.co>

CNSC. (s. f.). Estructura orgánica. Comisión Nacional del Servicio Civil. Recuperado el 27 de abril de 2026, de <https://www.cnsc.gov.co/transparencia/informacion-de-la-entidad/estructura-organica>

CNSC. (2025a). Acuerdo No. 21 de 2025: por el cual se convoca y se establecen las reglas del Proceso de Selección de Ingreso y Ascenso para proveer empleos en vacancia definitiva pertenecientes al Sistema Específico de Carrera Administrativa de la planta de personal de la Unidad Administrativa Especial Dirección de Impuestos y Aduanas Nacionales – DIAN, Proceso de Selección DIAN 2676. Comisión Nacional del Servicio Civil. <https://www.cnsc.gov.co/convocatorias/dian-2676>

CNSC. (2025b). Anexo Técnico del Acuerdo No. 21 de 2025: especificaciones técnicas de las diferentes fases del Proceso de Selección DIAN 2676. Comisión Nacional del Servicio Civil. <https://www.cnsc.gov.co/convocatorias/dian-2676>

CNSC. (2025c). Reporte de control de términos – Proceso de Selección DIAN – Grupo de Gerencia de Convocatoria 12: periodo 20 de agosto al 20 de octubre de 2025 [Registro interno de gestión documental, sistema de radicación PQRSJ]. Comisión Nacional del Servicio Civil.

Congreso de la República. (2004). Ley 909 de 2004. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=14861>

Congreso de la República. (2015). Ley 1755 de 2015: por medio de la cual se regula el derecho fundamental de petición y se sustituye un título del Código de Procedimiento Administrativo y de lo Contencioso Administrativo. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=65334>

DAFP. (2023). Modelo Integrado de Planeación y Gestión (MIPG). Departamento Administrativo de la Función Pública. <https://www.funcionpublica.gov.co/web/mipg>

DAMA International. (2017). DAMA-DMBOK2: Data management body of knowledge (2.^a ed.). Technics Publications.

Doshi-Velez, F. y Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv:1702.08608. <https://arxiv.org/abs/1702.08608>

Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2014). Metodología de la investigación (6.^a ed.). McGraw-Hill.

ISO. (2016). ISO 15489-1:2016: Information and documentation – Records management – Part 1: Concepts and principles. International Organization for Standardization. <https://www.iso.org/standard/62542.html>

ISO/IEC. (2022). ISO/IEC 27001:2022: Information security, cybersecurity and privacy protection – Information security management systems – Requirements. <https://www.iso.org/standard/82875.html>

Jiang, Y., Pang, P. C.-I., Wong, D. y Kan, H. Y. (2023). Natural language processing adoption in governments and future research directions: A systematic review. *Applied Sciences*, 13(22), 12346. <https://doi.org/10.3390/app132212346>

Johnson, J., Douze, M. y Jégou, H. (2019). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3), 535–547. <https://doi.org/10.1109/TBDATA.2019.2921572>

Jurafsky, D. y Martin, J. H. (s. f.). *Speech and language processing* (3.^a ed., borrador en línea, trabajo en preparación). Recuperado el 27 de abril de 2026, de <https://web.stanford.edu/~jurafsky/slp3/>

Klesel, M. y Wittmann, G. (2025). Retrieval-augmented generation (RAG). *Business & Information Systems Engineering*, 67, 551–561. <https://doi.org/10.1007/s12599-025-00945-3>

Lai, J., Gan, W., Wu, J., Qi, Z. y Yu, P. S. (2024). Large language models in law: A survey. *AI Open*, 5, 181–196. <https://doi.org/10.1016/j.aiopen.2024.09.002>

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. y Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>

Manning, C. D., Raghavan, P. y Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809071>

Microsoft. (2026). Precios de máquinas virtuales con GPU (serie NVads A10 v5). Microsoft Azure. <https://azure.microsoft.com/es-es/pricing/details/virtual-machines/>

MinTIC. (2019). Marco de interoperabilidad para Gobierno Digital. Ministerio de Tecnologías de la Información y las Comunicaciones. https://lenguaje.mintic.gov.co/sites/default/files/archivos/marco_de_interoperabilidad_para_gobierno_digital.pdf

MinTIC. (2022). Política de Gobierno Digital (Decreto 767 de 2022). Ministerio de Tecnologías de la Información y las Comunicaciones. <https://www.mintic.gov.co>

Neumann, O., Guirguis, K. y Steiner, R. (2022). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review*, 26(1), 114–141. <https://doi.org/10.1080/14719037.2022.2048685>

NIST. (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. <https://www.nist.gov/itl/ai-risk-management-framework>

OCDE. (2019). OECD principles on artificial intelligence. Organisation for Economic Co-operation and Development. <https://oecd.ai/en/ai-principles>

OCDE. (2025). AI in justice administration and access to justice. En Governing with artificial intelligence. Organisation for Economic Co-operation and Development. <https://www.oecd.org>

Oracle. (2026). Cloud price list: Compute, GPU. Oracle Cloud Infrastructure. <https://www.oracle.com/cloud/price-list/>

Pipitone, N. y Houir Alami, G. (2024). LegalBench-RAG: A benchmark for retrieval-augmented generation in the legal domain. arXiv:2408.10343. <https://arxiv.org/abs/2408.10343>

Presidencia de la República. (2015). Decreto 1083 de 2015. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=62866>

Sele, D. y Chugunova, M. (2024). Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. PLOS ONE, 19(2), e0298037. <https://doi.org/10.1371/journal.pone.0298037>

Senado de la República. (2022). Instructivo técnico para la elaboración de comunicaciones oficiales. <https://www.senado.gov.co>

Straub, V. J., van Noordt, C. y Zuiderwijk, A. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. Government Information Quarterly, 40(4), 101881. <https://doi.org/10.1016/j.giq.2023.101881>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. y Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>

Lista de Anexos

Los anexos cumplen una función de soporte documental y trazabilidad. Todo el contenido necesario para comprender el trabajo se encuentra en el cuerpo del documento; los anexos amplían la información y se referencian desde el texto cuando corresponde, indicando al lector el anexo específico al que puede dirigirse para profundizar.

Anexo	Nombre
A	Carta de autorización para la intervención empresarial
B	Datos de peticiones por etapa del proceso de selección DIAN 2022
C	Reporte de control de términos (período 250820 al 251020)
D	Cuestionario de percepción del Agente de IA
E	Encuesta de evaluación del Agente de IA (tipo Likert)
F	Repositorio del código fuente del agente

Anexo A - Carta de autorización para la intervención empresarial

A continuación, se presenta la carta de autorización emitida por la Comisión Nacional del Servicio Civil (CNSC), mediante la cual la entidad otorgó el aval institucional para el desarrollo del presente proyecto de grado en la modalidad de Intervención Empresarial, conforme a los lineamientos de la Universidad Ean.

Bogotá, 26 de septiembre de 2025

Señores,
Subproceso de Trabajos de Grado
Universidad Ean
Bogotá

Respetados señores,

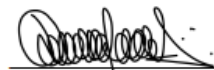
DIANA CAROLINA FIGUEROA MERIÑO, identificada con cédula de ciudadanía No. 45.556.643, me permito autorizar a **Juan Sebastián Ballen Villalba**, identificado con cédula de ciudadanía **1.192.757.676**, estudiante del programa de maestría **Ciencia de Datos** de la Universidad Ean, para que realice en nuestra entidad **Comisión Nacional del Servicio Civil (CNSC)**, su trabajo de grado titulado: **Agente de inteligencia artificial para la atención de solicitudes en la etapa de inscripciones de la CNSC**, bajo la modalidad de trabajo dirigido.

Nos comprometemos a brindar la información requerida para adelantar este proyecto académico, así mismo, autorizamos que el documento resultado de este trabajo sea publicado en el repositorio documental Minerva de la Universidad Ean.

A continuación, relacionamos los datos de la persona que será el contacto designado por la CNSC.

Nombre: Diana Carolina Figueroa Meriño
Cargo: Asesora Proceso de Selección
Celular: 3017602355
Correo: dfigueroa@cns.gov.co

Cordialmente,



DIANA CAROLINA FIGUEROA MERIÑO
CC 45.556.643

Anexo B - Datos de peticiones por etapa del proceso de selección DIAN 2022

El presente anexo contiene el registro histórico de peticiones ciudadanas recibidas y gestionadas durante el proceso de selección por mérito para la Dirección de Impuestos y Aduanas Nacionales (DIAN) correspondiente al año 2022. Los datos se organizan por etapa del proceso (inscripciones, verificación de requisitos mínimos, aplicación de pruebas y resultados) e incluyen el volumen de peticiones recibidas en cada fase. Esta información sirvió como insumo para contextualizar la magnitud operativa del proceso de atención de peticiones en la CNSC y fundamentar la pertinencia de la solución propuesta. La información fue obtenida de los registros internos de la entidad y empleada exclusivamente con fines académicos, garantizando el tratamiento responsable de los datos conforme a la Ley 1581 de 2012.

Dado el volumen de registros, en la tabla siguiente se presenta una muestra representativa de los datos. La tabla completa se encuentra disponible en el archivo fuente (Peticiones proceso de selección DIAN 2022.xlsx).

RADICADO ENTRADA	FECHA RADICADO	ETAPA	CONVOCATORIA	RADICADO DE SALIDA	FECHA RESPUESTA
2023RE000909	4/01/2023	PLANEACIÓN	DIAN 2022	2023RS002433	23/01/2023
2023RE008929	18/01/2023	VRM	DIAN 2022	2023RS006884	6/02/2023
2023RE012350	24/01/2023	PLANEACIÓN	DIAN 2022	2023RS006887	6/02/2023
2023RE026778	10/02/2023	INSCRIPCIONES	DIAN 2022	2023RS012357	20/02/2023
2023RE032273	16/02/2023	INSCRIPCIONES	DIAN 2022	2023RS015318	27/02/2023
2023RE031235	16/02/2023	INSCRIPCIONES	DIAN 2022	2023RS015203	27/02/2023
2023RE032800	16/02/2023	EXÁMENES MÉDICOS	DIAN 2022	2023RS016188	28/02/2023
2023RE031219	16/02/2023	INSCRIPCIONES	DIAN 2022	2023RS015829 - 2023RS016186	28/02/2023
2023RE032791	16/02/2023	INSCRIPCIONES	DIAN 2022	2023RS018752	3/03/2023
...	...	...	...	...	...

Nota. Elaboración propia a partir de registros internos de la CNSC (2022). Archivo fuente: Peticiones proceso de selección DIAN 2022.xlsx.

Anexo C - Reporte de control de términos (período 250820 al 251020)

El presente anexo contiene el reporte de control de términos correspondiente al período comprendido entre el 20 de agosto y el 20 de octubre de 2025, que constituyó el período de referencia para la estimación de la línea base operativa del proceso de atención de peticiones durante la etapa de inscripciones del proceso de selección DIAN. Los datos incluyen información por radicado sobre fechas de recepción, asignación, elaboración, revisión jurídica y envío de cada petición, permitiendo calcular los tiempos de ciclo por hito presentados en la Tabla 2 del documento principal. La información fue obtenida de los registros internos de la CNSC y empleada exclusivamente con fines académicos, garantizando el tratamiento responsable de los datos conforme a la Ley 1581 de 2012.

Dado el volumen de columnas del reporte original, en la tabla siguiente se presentan únicamente las columnas más relevantes para el análisis de tiempos de ciclo. El archivo completo se encuentra disponible en el archivo fuente (Reporte control términos 250820 al 251020.xlsx).

Número radicado de entrada (RE)	Fecha radicación (RE)	Fecha entrega Enrutador	Fecha producción	Fecha revisión	Fecha firma	Fecha de radicación (RS)	...
2025RE173141	20/08/2025	20/08/2025	20/08/2025	22/08/2025	28/08/2025	28/08/2025	...
2025RE173197	20/08/2025	20/08/2025	20/08/2025	25/08/2025	28/08/2025	28/08/2025	...
2025RE173358	20/08/2025	20/08/2025	26/08/2025	1/09/2025	3/09/2025	3/09/2025	...
2025RE173565	20/08/2025	20/08/2025	28/08/2025	2/09/2025	3/09/2025	3/09/2025	...
2025RE173956	20/08/2025	20/08/2025	29/08/2025	2/09/2025	5/09/2025	5/09/2025	...
...	...	...	...	...	...	...	...

Nota. Elaboración propia a partir de registros internos de la CNSC (2025). Archivo fuente: Reporte control términos 250820 al 251020.xlsx.

Anexo D - Cuestionario de percepción del Agente de IA

El presente anexo contiene el cuestionario de percepción aplicado a los funcionarios participantes del piloto controlado. El instrumento evalúa la utilidad percibida del agente, la calidad de los borradores generados y la disposición al uso de la herramienta en el flujo institucional de respuesta a peticiones ciudadanas, disponible en <https://forms.office.com/Pages/ResponsePage.aspx?id=WbVvwGgbhEuhT0fQ2DelqwQbPbqXM8RFvdW7KcAxRoRUN0kxNUIQRFJWQ1NDTjI2QUlyVDhZQ0FTUi4u>

Para asegurar la trazabilidad y la replicabilidad del estudio con independencia de la disponibilidad del formulario en línea, a continuación se transcribe el enunciado íntegro de las 13 preguntas abiertas del cuestionario de percepción:

1. ¿Podría describirme cómo fue su experiencia al utilizar el agente por primera vez?
¿Qué impresión le causó?
2. ¿Cuándo el agente le presentó un borrador de respuesta, que fue lo primero que reviso? ¿Qué tan útil le resultó como punto de partida?
3. ¿Como percibió la calidad de las citas normativas que el agente incluyó en los borradores? ¿Eran pertinentes, precisas, verificables?
4. ¿Cuándo una petición contenía varias preguntas o solicitudes, como manejo el agente esa situación? ¿Le pareció adecuado?
5. ¿Considera que el uso del agente redujo el tiempo que usted dedica a elaborar un borrador de respuesta? ¿En que etapas del proceso percibió la mayor diferencia?
6. ¿El uso del agente tuvo algún efecto en la calidad final de las respuestas?
7. En comparación con el proceso manual, ¿qué aspectos mejoraron con el uso del agente y cuales no cambiaron o incluso empeoraron?
8. ¿Cuáles fueron las principales limitaciones o dificultades que encontró al usar el agente durante el piloto?
9. ¿Qué funcionalidades o mejoras le gustaría que tuviera el agente en una versión futura?
10. Desde su experiencia, ¿qué barreras podrían dificultar que el equipo adopte el agente de forma permanente en el proceso?
11. Si tuviera que calificar su experiencia general con el agente en una escala de 1 a 10, ¿qué nota le daría y por qué?

12. ¿Recomendaría el uso del agente a otros equipos de la CNSC o a otros procesos de selección? ¿Por qué?

13. ¿Hay algo más que quiera agregar sobre su experiencia con el agente que no hayamos abordado en este formulario?

Anexo E - Encuesta de evaluación del Agente de IA (tipo Likert)

El presente anexo contiene la encuesta estructurada tipo Likert (escala de 1 a 5) aplicada a los funcionarios que elaboran respuestas a peticiones en el proceso de selección DIAN. El instrumento está compuesto por ítems organizados en las siete dimensiones del modelo diagnóstico de madurez, disponible en <https://forms.office.com/Pages/ResponsePage.aspx?id=WbVvwGgbhEuhT0fQ2DelqwQbPbqXM8RFvdW7KcAxRoRUNzY3UDcwTkVaR0IIMzZBRFVSMIJXMDQ2Qy4u>

A continuación se transcribe el enunciado íntegro de los 25 ítems de la encuesta tipo Likert (escala de 1 = totalmente en desacuerdo a 5 = totalmente de acuerdo), organizados por las siete dimensiones del instrumento:

Dimensión 1, Apoyo en la identificación normativa y estructuración de la respuesta

1. El agente me ayudo a identificar la normativa aplicable a la petición de forma mas rapida que el proceso manual.
2. Los borradores generados por el agente constituyeron un punto de partida util para la elaboracion de la respuesta final.
3. El agente facilito la estructuracion de la respuesta (saludo, antecedentes, fundamento juridico, decision, cierre).
4. Considero que el agente reduce la dependencia en la experiencia individual del funcionario para elaborar respuestas.

Dimensión 2, Calidad y precisión de las citas normativas

5. Las citas normativas incluidas en los borradores del agente fueron pertinentes para la petición consultada.
6. Las referencias a artículos, numerales y fuentes normativas generadas por el agente fueron precisas y verificables.
7. El agente cito normativa vigente y aplicable al proceso de seleccion DIAN 2676 de 2025.
8. Las evidencias normativas presentadas por el agente me permitieron verificar la fundamentacion juridica del borrador.

Dimensión 3, Eficiencia y reducción de tiempos

9. El uso del agente redujo el tiempo que normalmente dedico a buscar normativa aplicable.

10. El tiempo total para elaborar un borrador de respuesta fue menor al utilizar el agente como apoyo.

11. El agente me permitio dedicar mas tiempo a la revision de fondo y menos a la redaccion inicial del borrador.

Dimensión 4, Calidad de redacción y estructura argumentativa

12. La estructura argumentativa de los borradores fue coherente y logica.

13. El lenguaje utilizado por el agente en los borradores fue adecuado para una comunicacion oficial de la CNSC.

14. Los borradores del agente requirieron pocos ajustes antes de ser enviados a revision juridica.

Dimensión 5, Comprensión e interpretación de la petición

15. El agente identifico correctamente la(s) solicitud(es) del ciudadano contenida(s) en la peticion.

16. Cuando la peticion contenia multiples preguntas, el agente las segmento y abordo de manera individual.

17. La interpretacion que el agente hizo de la solicitud del ciudadano fue acorde con lo que el peticionario solicitaba.

Dimensión 6, Experiencia de usuario (UX) e interfaz

18. La interfaz del agente fue intuitiva y facil de utilizar.

19. Las instrucciones proporcionadas durante la capacitacion fueron suficientes para operar el agente.

20. El tiempo de respuesta del agente (velocidad de generacion del borrador) fue aceptable.

21. Los resultados del agente (borrador, evidencias, evaluacion) se presentaron de forma clara y organizada.

Dimensión 7, Confianza, revisión humana y disposición de adopción

22. Me sentí seguro(a) utilizando los borradores del agente como base para la respuesta final, sin que esto reemplazara mi revisión profesional antes del envío

23. La información generada por el agente fue lo suficientemente confiable para apoyar mis decisiones durante la elaboración de la respuesta.

24. Estaría dispuesto(a) a utilizar el agente de forma permanente y a recomendarlo como herramienta de apoyo en otros equipos o procesos de la CNSC.

25. Considero que el agente tiene potencial para mejorar la calidad de las respuestas a peticiones e incorporar funcionalidades adicionales en el futuro.

Anexo F - Repositorio del código fuente del agente

El código fuente completo del agente se encuentra disponible en el siguiente repositorio público de GitHub: <https://github.com/sebastiaanbv/Asistente-peticiones>

Para asegurar la reproducibilidad de los resultados reportados, el estado del agente evaluado en el piloto (13 al 30 de abril de 2026) corresponde a la versión consolidada del prototipo en la rama principal del repositorio (commit dafebdb, MVP), que integra el pipeline RAG híbrido, el segmentador de peticiones, el evaluador de borradores y el formateador documental empleados durante el piloto. El repositorio fue saneado para su publicación: se reescribió el historial con mensajes en formato convencional, se retiraron del control de versiones los datos personales de los aspirantes, que permanecen en almacenamiento local conforme a la Ley 1581 de 2012, y se declararon la licencia AGPL-3.0-or-later, una política de seguridad (SECURITY.md) y la autoría del proyecto (AUTHORS). Se recomienda asociar a la versión del piloto una etiqueta específica (por ejemplo, v1.0-piloto) y archivar el repositorio en un repositorio de datos con identificador persistente (DOI), como Zenodo, para su citación académica formal.

La Tabla presenta la estructura de directorios y los módulos principales del repositorio con su descripción funcional.

Ruta / Archivo	Descripción
src/	Código fuente principal del agente
app_streamlit.py	Interfaz web (Streamlit); punto de entrada del usuario
rag_pipeline.py	Orquestador del pipeline RAG completo
rag_core.py	Búsqueda híbrida FAISS + BM25 con RRF y reranking
petition_segmenter.py	Segmentador multi-pregunta de peticiones ciudadanas
generate_response.py	Generación del borrador de respuesta con el LLM
draft_evaluator.py	Evaluador automático de calidad del borrador (A/B/C)
fill_docx.py	Exportación del borrador al formato Word institucional
prompting.py	Construcción de prompts y control de citas normativas
norm_ref.py	Validación y extracción de referencias normativas
llm_client.py	Cliente de comunicación con Ollama (qwen2.5:3b)
build_index.py	Construcción del índice FAISS y BM25 desde el corpus
config.py	Carga y resolución de parámetros de configuración

Ruta / Archivo	Descripción
docx_formatter.py	Formateador estructurado de documentos Word
configs/	Archivos de configuración del sistema
settings.yaml	Parámetros del pipeline: modelos, rutas, umbrales
templates/	Plantillas de instrucción y formato de salida
respuesta_base.md	Modelo base de respuesta institucional CNSC
format_only.md	Instrucciones de formato para el LLM
corpus/	Repositorio normativo (acuerdos, leyes, decretos)
indexes/	Índices FAISS y BM25 generados desde el corpus
docs/	Documentación técnica y bitácora del proyecto