Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y Tecnología

Máster Universitario en Dirección y Gestión de Tecnologías de la Información

Propuesta de Framework de Gobernanza para la Adopción Responsable de Inteligencia Artificial

Aplicado a Entidades del Sector Público del Distrito Capital de Bogotá

Trabajo fin de estudio presentado por:	Mario Alexander Ortiz Salgado Javier Mauricio Rocha Cruz Maria Alejandra Rodriguez Salas
Tipo de trabajo:	TFM
Director/a:	José Ramón Coz Fernández
Fecha:	28 de enero 2026

Resumen

Este Trabajo Fin de Máster surge ante la necesidad de las entidades del sector público de Bogotá de integrar la inteligencia artificial (IA) bajo estándares éticos y legales, abordando la brecha existente entre el marco normativo nacional y las capacidades operativas reales del distrito. El objetivo principal de esta investigación es diseñar, desarrollar y validar una propuesta de Framework de Gobernanza de IA adaptado al contexto distrital, que permita una adopción responsable, transparente y centrada en el ciudadano, facilitando el tránsito de la teoría normativa a la ejecución operativa.

La metodología empleada se basa en el enfoque Design Science Research (DSR), estructurado en fases de diagnóstico, diseño de artefactos y validación. El estudio incluyó un análisis comparativo de marcos internacionales y locales, seguido del desarrollo de un Toolkit operativo que incorpora herramientas de evaluación de impacto y matrices de riesgos adaptadas a la realidad institucional.

Los resultados demuestran que la implementación de una estructura de gobernanza organizada por capas permite la identificación temprana de sesgos y fallos de seguridad. Como conclusión, se establece que el framework propuesto no solo reduce la incertidumbre jurídica e institucional en las entidades de Bogotá, sino que garantiza que la innovación tecnológica fortalezca la confianza pública y la eficiencia administrativa sin comprometer los derechos fundamentales de la población.

Como líneas de trabajo futuro, se propone escalar la aplicación del framework a otros sectores administrativos del distrito como salud y movilidad, además de diseñar indicadores de madurez institucional para monitorear el progreso de las entidades. Asimismo, se plantea la expansión del modelo a niveles de gobierno nacionales y su armonización con regulaciones internacionales emergentes como el AI Act de la Unión Europea.

Palabras clave: Gobernanza de IA, Sector Público, Framework de Adopción, Ética Algorítmica, Gestión de Riesgos.

Abstract

This Master's Thesis arises from the need for public sector entities in Bogotá to integrate artificial intelligence (AI) under ethical and legal standards, addressing the existing gap between the national regulatory framework and the district's real operational capabilities. The main objective of this research is to design, develop, and validate a proposed AI Governance Framework adapted to the district context, which allows for a responsible, transparent, and citizen-centered adoption, facilitating the transition from regulatory theory to operational execution.

The methodology employed is based on the Design Science Research (DSR) approach, structured into phases of diagnosis, artifact design, and validation. The study included a comparative analysis of international and local frameworks, followed by the development of an operational Toolkit that incorporates impact assessment tools and risk matrices adapted to the institutional reality.

The results demonstrate that the implementation of a governance structure organized by layers allows for the early identification of biases and security failures. In conclusion, it is established that the proposed framework not only reduces legal and institutional uncertainty in Bogotá's entities but also ensures that technological innovation strengthens public trust and administrative efficiency without compromising the fundamental rights of the population.

As future lines of work, it is proposed to scale the application of the framework to other administrative sectors of the district such as health and mobility, in addition to designing institutional maturity indicators to monitor the progress of the entities. Likewise, the expansion of the model to national government levels and its harmonization with emerging international regulations such as the European Union's AI Act is proposed.

Keywords: AI Governance, Public Sector, Adoption Framework, Algorithmic Ethics, Risk Management.

Índice de contenidos

Índice de figuras

Índice de tablas

Organización del trabajo en grupo

La organización del trabajo del presente capítulo se estructuró mediante un enfoque metodológico orientado a garantizar la coherencia entre los objetivos específicos del proyecto, las actividades desarrolladas y los entregables generados. Para tal fin, se estableció un flujo de trabajo sistemático que define con precisión las fases del desarrollo del framework, los responsables directos de cada actividad y los artefactos finales producidos en cada etapa. Este enfoque permite asegurar trazabilidad, claridad operativa y alineación con los estándares de investigación académica y con las necesidades reales de las entidades públicas del Distrito de Bogotá.

Los mecanismos de coordinación implementados para llevar a cabo la distribución de tareas del equipo aplicados son: Identificación de backlog de secciones, asignación de cada sección y sprints semanales para priorización, desarrollo y revisión de los entregables realizando procesos de validación e inspección de calidad del entregable por parte de los responsables de apoyo. De esta manera se garantizó que el trabajo ha sido entendido, verificado y aprobado por todo el equipo.

Como parte de la investigación de los marcos internacionales y locales se asignó de manera equilibrada la profundización y especialización de cada documento, y se llevaron a cabo sesiones de socialización durante la fase del desarrollo del estado del arte, garantizando que las fuentes bibliográficas y la información relevante del proyecto haya sido entendida por todos los integrantes del equipo.

La Tabla 1 presenta el flujo de trabajo consolidado, detallando las etapas del proceso, los roles implicados, los insumos requeridos y los productos finales generados. Esta estructura permitió organizar de manera eficiente la elaboración del capítulo 4, que incluye desde la definición del marco conceptual de gobernanza de IA hasta el desarrollo del toolkit operativo y sus mecanismos de validación.

Tabla 1. Organización del trabajo en grupo.

Capítulo / Sección	*Responsable Principal*	*Responsables de Apoyo*	*Tareas Clave y Justificación*
CAPÍTULO 1: INTRODUCCIÓN	Javier Mauricio	Mario, María	Liderar la integración y pulir la narrativa global. Asegurar que el capítulo sirva como un marco claro.
1.1 Contexto y Planteamiento del Problema	Javier Mauricio	Mario	Reforzar el "storytelling". Transformar la descripción en una narrativa convincente.
1.2 Justificación	Mario Alexander	María	Enfatizar la relevancia práctica y política. Revisar los tres ejes de la brecha.

Capítulo / Sección	*Responsable Principal*	*Responsables de Apoyo*	*Tareas Clave y Justificación*
1.3 Planteamiento de la Solución	Javier Mauricio	Mario	Articular con precisión técnica el artefacto (framework de 5 capas y el toolkit).
1.4 Estructura del Trabajo	María Alejandra		Actualizar y asegurar consistencia. Verificar que la descripción de los capítulos se ajuste al contenido final.
CAPÍTULO 2: ESTADO DEL ARTE	Maria Alejandra	Javier, Mario	Liderar la consolidación técnica y la síntesis. Asegurar un análisis crítico.
2.1 Introducción General	María Alejandra	Javier	Sintetizar y enfocar. Reescribir para preparar al lector para el análisis comparativo.
2.2 Marcos Internacionales de Gobernanza de IA	Mario Alexander	Maria	Sistematizar y comparar. Consolidar la información de los marcos (UE, NIST, OCDE, etc.).
2.3 Estudios Académicos Relevantes	María Alejandra	Javier	Tejer una narrativa académica. Ir más allá de resumir estudios individuales.
2.4 Casos de Implementación en América Latina	Mario Alexander	María	Extraer lecciones aprendidas. Reorganizar para destacar hallazgos comunes y vincular a Bogotá.
2.5 Metodología de Revisión Sistemática	Javier Mauricio	María	Robustecer la sección metodológica. Asegurar rigor y transparencia. Dar formato a la Tabla 1.
2.6 Análisis Comparativo y Brechas de Implementación	Javier Mauricio	Mario, María	Conducir el análisis que sintetiza todo lo anterior. Responsable clave de la Tabla 2.
2.7 Contribución Esperada y Originalidad	Mario Alexander	Javier, María	Vender la propuesta. Definir de manera clara y convincente la brecha que llena el TFM.
CAPÍTULO 3: OBJETIVOS Y METODOLOGÍA	Mario Alexander	Javier, María	Liderar la claridad y viabilidad. Asegurar que objetivos sean SMART y la metodología DSR sea reproducible.
3.1 Objetivo Principal	Mario Alexander		Precisar y alinear con el problema (Cap. 1) y la brecha (Cap. 2).
3.2 Objetivos Específicos	Mario Alexander	Javier, María	Asegurar la trazabilidad. Verificar que cubran todo el

Capítulo / Sección	*Responsable Principal*	*Responsables de Apoyo*	*Tareas Clave y Justificación*
			ciclo de desarrollo del framework.
3.3 Metodología de Trabajo (Fases DSR)	Mario Alexander	Javier, María	Operacionalizar el DSR. Detallar cada fase con actividades e instrumentos de validación claros.
CAPÍTULO 4: DESARROLLO DE LA PROPUESTA Y ANÁLISIS DE RESULTADOS	Equipo		Integrar las contribuciones de todos, garantizar coherencia narrativa y estilo unificado.
4.1 Arquitectura del Framework de Gobernanza de IA	Mario Alexander	Mario, María	Reescribir con un enfoque técnico-arquitectónico. Crear o refinar diagramas.
4.2 Caja de Herramientas Operativa (Toolkit)	Javier Mauricio	Mario, María	Liderar el refinamiento de todas las plantillas. Asegurar consistencia y usabilidad autónoma.
4.3 Análisis de Resultados de Validación	Mario Alexander	Maria, Javier	Liderar la reescritura con rigor metodológico. Estructurar la sección de la validación.
4.3.1 Selección de casos de validación	Javier Mauricio	Mario alexander	Simular los resultados con claridad académica. Analizar cualitativamente los comentarios.
4.3.2 Validación Experimental mediante Simulación	Maria Alejandra	Javier, Mario	Narrar los casos de estudio. Evidenciar desafíos, proceso y resultados, destacando el valor del framework.
4.3.3 Hallazgos Emergentes y Ajustes Finales	Maria Alejandra		Sintetizar y contrastar. Extraer lecciones aprendidas transversales.
4.3.4 Hallazgos Emergentes y Ajustes Finales	Javier Mauricio	Javier, Mario	Documentar la evolución del framework. Listar claramente los ajustes realizados por la validación.
4.4 Síntesis del Capítulo	Mario Alexander	Maria, Javier	Redactar una conclusión poderosa. Resumir hallazgos clave y cómo el framework responde a los objetivos.
CAPÍTULO 5: CONCLUSIONES	Mario Alexander	María, Javier	Liderar la visión estratégica y de aporte. Centrarse en el impacto y el legado de la investigación.
5.1 Conclusiones	Maria Alejandra	Mario	Sistematizar y evidenciar el cumplimiento. Demostrar

Capítulo / Sección	*Responsable Principal*	*Responsables de Apoyo*	*Tareas Clave y Justificación*
			cómo se respondió a cada objetivo.
5.2 Líneas de Trabajo Futuro	Javier Mauricio	María	Proyectar la investigación hacia adelante. Plantear caminos concretos para escalar y mejorar la propuesta.

Fuente: Elaboración propia.

Una vez definido el flujo de trabajo, cada fase se desarrolló de acuerdo con los siguientes lineamientos:

Coherencia metodológica: Todas las actividades están directamente vinculadas con los objetivos específicos establecidos en el Capítulo 3, asegurando consistencia y evitando desviaciones conceptuales o metodológicas.

Enfoque en aplicabilidad institucional: Las tareas fueron definidas tomando como referencia la estructura organizacional, la normativa vigente y la capacidad operativa de las entidades públicas distritales, con el fin de asegurar que los resultados puedan ser implementados en la práctica.

Trazabilidad y control de calidad: Cada entregable dispone de criterios de revisión y validación (internos y externos), lo que garantiza rigor técnico y pertinencia para el sector público.

Gestión colaborativa: La asignación de roles dentro del flujo permitió una división clara del trabajo entre los integrantes del equipo, favoreciendo la especialización y la calidad de cada componente.

1. Introducción

La inteligencia artificial (IA) ha emergido como una fuerza transformadora en el sector público, con el potencial de redefinir los fundamentos mismos de la gestión y la prestación de servicios. Sin embargo, su adopción conlleva riesgos significativos en materia de derechos fundamentales, equidad, transparencia y seguridad. A nivel internacional, se han desarrollado marcos de gobernanza robustos para guiar una implementación ética y efectiva, como el AI Act de la Unión Europea, los Principios de la OCDE y el NIST AI RMF. No obstante, existe una brecha crítica entre la existencia de estos marcos normativos y su aplicabilidad práctica en contextos subnacionales con capacidades institucionales y técnicas heterogéneas.

Este trabajo se centra en el Distrito Capital de Bogotá, donde la desconexión entre el marco nacional (Consejo Nacional de Política Económica y Social [CONPES], 2024) y las realidades operativas de las entidades distritales obstaculiza la adopción responsable de la IA. El problema central radica en la falta de un framework de gobernanza contextualizado que traduzca los principios abstractos en herramientas operativas viables. Para abordar esta brecha, esta investigación propone el diseño, desarrollo y validación de un Framework de Gobernanza de IA integral y práctico, junto con una caja de herramientas asociada, específicamente adaptado al sector público bogotano. La contribución principal será un artefacto validado empíricamente que permita a las entidades distritales adoptar sistemas de IA de manera ética, segura, legal y efectiva, superando así las limitaciones de adaptación contextual, fragmentación implementativa y déficit de validación empírica identificadas en la literatura.

1.1.Justificación

El problema específico que aborda este trabajo es la brecha de implementación entre el marco normativo nacional de IA (CONPES, 2024), y las capacidades operativas reales de las entidades públicas del Distrito Capital de Bogotá. Esta brecha se manifiesta en tres limitaciones convergentes, identificadas a partir del análisis del estado del arte. En primer lugar, la falta de adaptación contextual: los marcos internacionales y nacionales existentes no se ajustan adecuadamente a las particularidades institucionales, técnicas y políticas del Distrito Capital, como las capacidades heterogéneas entre entidades, la dependencia de proveedores tecnológicos externos y la inherente inestabilidad política de las administraciones subnacionales (OECD, 2024). En segundo término, la fragmentación implementativa: no existe una integración sistemática entre las dimensiones críticas de la gobernanza de IA (principios éticos, requisitos normativos, controles técnicos, herramientas operativas y métricas de evaluación), lo que genera esfuerzos dispersos e incoherentes. Finalmente, el déficit de validación

empírica: los marcos disponibles carecen de evidencia robusta sobre su efectividad en contextos subnacionales reales, lo que restringe su aplicabilidad práctica y viabilidad operativa (Departamento Nacional de Planeación, 2025).

Este problema es de suma relevancia para la comunidad científica y de políticas públicas. La investigación académica ha demostrado exponencialmente el potencial de la IA para mejorar la eficiencia, la transparencia y la lucha contra la corrupción en el sector público (Uñate y Hernández, 2022; Cotino y Castellanos, 2023). Sin embargo persiste una laguna respecto a la traducción de principios abstractos y marcos nacionales en instrumentos prácticos y adaptados para gobiernos intermedios y locales. Estudios como los de Ruiz Sánchez (2024) sobre la adopción en las Altas Cortes colombianas y el análisis de Gutiérrez Peña (2024) sobre las políticas nacionales, confirman avances, pero también señalan falta de continuidad y mecanismos de seguimiento. La Carta Iberoamericana de IA (Revista CLAD, 2024) y los análisis comparativos regionales (González Méndez y Vásquez López, 2024) subrayan la heterogeneidad en madurez y la necesidad de enfoques contextualizados. Por lo tanto, este TFM pretende llenar este vacío desarrollando y validando un framework de gobernanza específico para el nivel subnacional, asegurando que la promesa transformadora de la IA se realice de manera responsable y efectiva en el contexto específico de Bogotá.

1.2. Planteamiento de la Solución

Para solucionar la brecha identificada, esta investigación propone el diseño, desarrollo y validación de un Framework de Gobernanza de IA integral y práctico, junto con su caja de herramientas asociada (toolkit). La solución general consiste en un artefacto estructurado en cinco capas interoperables: (1) Principios y políticas, (2) Modelo de gobierno, (3) Proceso de ciclo de vida, (4) Controles clave, y (5) Métricas/KPIs. Este diseño asegura la trazabilidad de la normativa nacional (CONPES, 2024) y los marcos internacionales de referencia (UE, OCDE, UNESCO, NIST, ISO), al mismo tiempo que se adapta flexiblemente a las capacidades reales del sector público distrital.

El objetivo general es permitir a las entidades públicas del Distrito Capital de Bogotá adoptar e implementar sistemas de inteligencia artificial de manera ética, segura, legal y efectiva. El enfoque o filosofía de la solución se basa en el paradigma de Design Science Research (DSR), el cual es idóneo para la creación y validación de artefactos que resuelven problemas organizacionales identificados (Hevner et al., 2004). La metodología implica un proceso iterativo que incluye el análisis contextual, el diseño del framework y toolkit, y su validación experimental a través de la simulación y pilotos controlados en servicios públicos distritales, midiendo el impacto en el cumplimiento normativo, la eficiencia del servicio y la gestión de riesgos.

1.3. Estructura del Trabajo

Este trabajo se estructura en los siguientes capítulos:

Capítulo 2: Estado del Arte. Se profundizará en el marco teórico y conceptual, presentando un análisis comparativo detallado de los marcos internacionales de gobernanza de IA (AI Act de la UE, NIST AI RMF, Principios de la OCDE, Recomendación de la UNESCO, estándares ISO) y una revisión sistemática de la literatura académica y casos de implementación relevantes en Colombia y América Latina. Esto sentará las bases para identificar las brechas y limitaciones que justifican la investigación.

Capítulo 3: Objetivos y Metodología. Se detallarán el objetivo principal y los objetivos específicos de la investigación. Asimismo, se describirá en profundidad la metodología de trabajo basada en el Design Science Research (DSR), desglosada en cuatro fases secuenciales: (1) Análisis Contextual y de Requisitos, (2) Diseño del Framework y Desarrollo del Toolkit, (3) Validación Experimental y Refinamiento, y (4) Consolidación y Elaboración de la Guía de Implementación. Se especificarán las actividades, instrumentos y técnicas de validación para cada fase.

Capítulo 4: Desarrollo de la Propuesta / Análisis de Resultados. Se presentará el framework de gobernanza de IA propuesto en su totalidad, desglosando su arquitectura de cinco capas y cada uno de los componentes de la caja de herramientas operativa (toolkit). También se expondrán los resultados y el análisis derivado del proceso de validación a través de la selección y simulación de un caso de uso demostrativo propio de la secretaría distrital de Bogotá.

Capítulo 5: Conclusiones. Se expondrán las principales conclusiones derivadas del trabajo, reflexionando sobre el grado de cumplimiento de los objetivos, la utilidad del framework desarrollado y sus limitaciones. Finalmente, se propondrán líneas futuras de investigación y recomendaciones para la implementación escalonada del framework en el Distrito Capital de Bogotá.

Anexos: Entregables. Esta sección referencia los entregables del trabajo que detallan los resultados de la investigación. Se incluye una tabla comparativa de frameworks derivada del análisis de marcos internacionales (Anexo A), la composición de la arquitectura del framework desarrollado (Anexo B), una vista del dashboard para la obtención de métricas y KPIs (Anexo C), el documento base de implementación del ciclo de vida de gobernanza de IA (Anexo D), los formatos obtenidos del ejercicio de simulación con cada herramienta (Anexo E) y, por último, los manuales de usuario correspondientes (Anexo F).

2. Contexto y estado del arte

2.1. Introducción General

La incorporación de la inteligencia artificial (IA) en el sector público va más allá de la simple innovación tecnológica para actuar como un impulsor que transforma las bases de la gestión pública. Si bien se dispone de marcos normativos internacionales ampliamente reconocidos, la sola existencia de estos no garantiza su efectiva implementación, una realidad que se ha reflexionado en el aspecto subnacional, como en el caso objeto de estudio para el Distrito Capital de Bogotá. Este capítulo tiene el propósito de delinear la matriz normativa y académica que sustentaría la exigencia de un framework de gobernanza para dicho contexto, evidenciando un camino que inicia desde las directrices globales hasta las realidades operacionales en el nivel local y las brechas estructurales que obstaculizan su aplicación práctica.

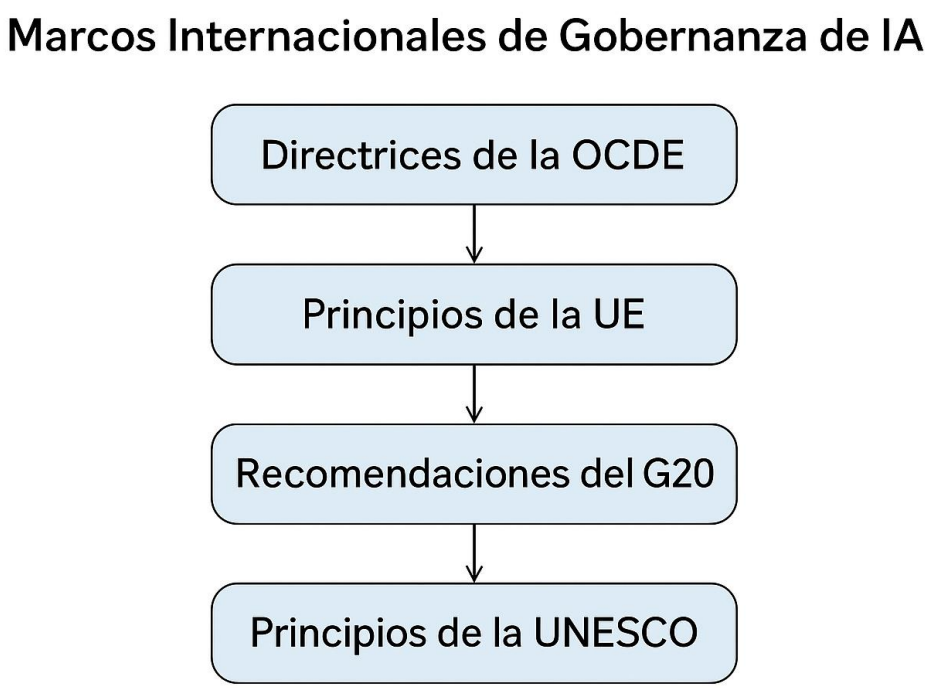
Es esencial destacar que la brecha tecnológica, la débil capacidad institucional y la dependencia de los proveedores externos crean un ambiente donde deben adaptarse significativamente los modelos globales antes que puedan ser utilizados en gobiernos locales. Se suma a estos desafíos la propia inestabilidad política inherente a las administraciones subnacionales, además de las importantes diferencias técnicas que existen entre comisiones distritales, estos son sólo algunos de factores que complican aún más el escenario.

2.2. Marcos Internacionales de Gobernanza de IA

La gobernanza de la inteligencia artificial en el sector público ha emergido como una prioridad estratégica a nivel global, generando el desarrollo de múltiples marcos regulatorios y estándares internacionales. El análisis comparativo de estos frameworks revela enfoques diversos pero complementarios para abordar los desafíos éticos, técnicos y regulatorios que plantea la adopción de IA en entidades gubernamentales.

La Figura 1 presenta una síntesis visual de los principales marcos internacionales de gobernanza de la inteligencia artificial, destacando cuatro referentes ampliamente adoptados a nivel global: las Directrices de la OCDE, los Principios de la Unión Europea, las Recomendaciones del G20 y los Principios Éticos de la UNESCO. Estos lineamientos conforman la base conceptual para el desarrollo de políticas públicas, estándares regulatorios y mecanismos de responsabilidad en el uso de IA, y sirven como referentes esenciales para la construcción de modelos de gobernanza en gobiernos nacionales y locales.

Figura 1 Marcos Internacionales de Gobernanza de IA.



Fuente: Elaboración propia (2025).

La Unión Europea (2024) desarrolló el marco regulatorio más integral mediante el AI Act 2024/1689, estableciendo un enfoque basado en riesgo que categoriza los sistemas de IA en cuatro niveles: sistemas prohibidos, alto riesgo, riesgo limitado y riesgo mínimo. Para el sector público, este marco resulta particularmente relevante en sistemas de evaluación y priorización de beneficiarios, biometría, chatbots ciudadanos y analítica de fraude social (European Union, 2024). El reglamento establece obligaciones específicas de gestión de riesgos, gobernanza de datos, documentación técnica y supervisión humana para entidades gubernamentales que implementen sistemas de alto riesgo.

El National Institute of Standards and Technology (2023) adoptó un enfoque técnico-operacional mediante el AI Risk Management Framework, centrado en cuatro funciones principales: gobernar, mapear, medir y gestionar riesgos. Este framework, complementado por el Perfil GenAI específico para sistemas generativos, proporciona metodologías prácticas para la identificación, evaluación y mitigación de riesgos a lo largo del ciclo de vida de los sistemas de IA (NIST, 2024). Su orientación técnica lo hace adaptable para organizaciones que buscan implementar procesos sistemáticos de gestión de riesgos sin las restricciones de un marco regulatorio obligatorio.

La Organisation for Economic Co-operation and Development (2024) estableció una base conceptual para el desarrollo de políticas nacionales mediante los Principios de IA actualizados, promoviendo IA

innovadora y confiable que respete los derechos humanos, valores democráticos, transparencia, explicabilidad, robustez y rendición de cuentas. Estos principios han influido significativamente en el diseño de la hoja de ruta nacional (CONPES, 2024), proporcionando un marco orientativo para la formulación de políticas públicas que equilibren la innovación tecnológica con la protección de derechos fundamentales (Departamento Nacional de Planeación, 2025).

En el ámbito nacional, la Ley 2279 (2022) complementa el marco regulatorio (CONPES, 2024) al establecer el marco habilitante para la transformación digital del Estado colombiano, incluyendo la implementación de tecnologías emergentes como la inteligencia artificial en la prestación de servicios públicos. El Congreso de la República avanza además en la discusión de una regulación específica de IA que armonizará los estándares internacionales con las particularidades del contexto colombiano.

UNESCO (2021) aportó una perspectiva centrada en valores humanos mediante la Recomendación sobre la Ética de la IA, estableciendo principios éticos fundamentales para el desarrollo y despliegue de sistemas de IA. Su enfoque en la equidad, no discriminación, transparencia y participación ciudadana resulta relevante para entidades públicas que manejan servicios esenciales y datos sensibles de la ciudadanía. Las orientaciones específicas para el uso responsable de IA generativa en administración pública proporcionan directrices prácticas para la implementación ética de estas tecnologías (UNESCO, 2023).

El G7 Digital and Technology Ministers (2023) estableció principios internacionales y un código de conducta voluntario para desarrolladores de IA mediante el Proceso de Hiroshima, enfocándose en la gestión de riesgos a lo largo del ciclo de vida, reporte transparente de capacidades y limitaciones, y compartición responsable de información. Aunque de naturaleza voluntaria, estos principios han influido en el desarrollo de marcos regulatorios nacionales y proporcionan estándares de referencia para la cooperación internacional en gobernanza de IA.

Los estándares ISO/IEC 23894 e ISO/IEC 42001 complementan estos marcos con enfoques técnicos específicos. La International Organization for Standardization (2023a) proporciona directrices para la gestión de riesgos de IA a lo largo del ciclo de vida mediante el ISO/IEC 23894, mientras que el ISO/IEC 42001 establece requisitos para sistemas de gestión de IA con enfoque de mejora continua (International Organization for Standardization, 2023b). Estos estándares resultan valiosos para entidades públicas que buscan certificación internacional e implementación de mejores prácticas técnicas.

La convergencia de estos marcos internacionales refleja un consenso emergente sobre la necesidad de enfoques sistemáticos, basados en riesgos y centrados en valores humanos para la gobernanza de IA en el sector público. Cada framework aporta perspectivas complementarias que pueden integrarse en

modelos nacionales adaptados a contextos institucionales específicos, como el propuesto para el Distrito Capital de Bogotá.

2.3. Estudios Académicos Relevantes

La investigación académica en gobernanza de IA para el sector público ha experimentado un crecimiento exponencial desde 2020, generando contribuciones significativas en áreas de evaluación de impacto, prevención de corrupción, transparencia algorítmica y modelos de gobernanza colaborativa.

El trabajo sobre la fusión transformadora entre el sector público y la inteligencia artificial propuso el Test de Evaluación de Impacto de la Inteligencia Artificial (TEI-Ai), un marco metodológico que permite a las organizaciones públicas evaluar el alcance y desafíos asociados con la implementación de IA (Vestri, 2024). Este instrumento esquemático aborda dimensiones críticas como datos, tecnología, recursos básicos, habilidades humanas y capacidades intangibles, proporcionando una herramienta práctica para la toma de decisiones informadas sobre adopción de IA en el sector público. El marco del TEI-Ai establece que cualquier entidad pública que introduzca IA debe construir una estrategia que considere diversos criterios derivados de esta tecnología disruptiva.

La investigación sobre inteligencia artificial contra la corrupción en el sector público examinó la convergencia entre ciencias del comportamiento e IA para el control preventivo de la corrupción (Uñate y Hernández, 2022). Este estudio identificó oportunidades significativas para mejorar los mecanismos de supervisión y control, pero también advirtió sobre riesgos específicos como sesgos algorítmicos, opacidad en la toma de decisiones y potenciales vulneraciones de derechos fundamentales. Los hallazgos sugieren la necesidad de marcos éticos robustos y mecanismos de rendición de cuentas específicos para aplicaciones anticorrupción.

Desde otra perspectiva, Cotino y Castellanos (2023) realizaron un análisis de algoritmos abiertos y que no discriminen en el sector público, abordando los desafíos éticos, legales y sociales que plantea la IA en la administración pública, con énfasis en la prevención de sesgos algorítmicos y la garantía de transparencia. La investigación propone soluciones como evaluaciones de impacto algorítmico, auditorías periódicas, supervisión humana y registros públicos, elementos que han sido incorporados en marcos regulatorios como el AI Act europeo y modelos prácticos como el canadiense. Los autores destacan la necesidad de sistemas de IA responsables y respetuosos con los derechos humanos, promoviendo una gobernanza algorítmica ética.

López (2021) desarrolló un análisis sobre la aplicación de IA en gestión pública, destacando cómo la implementación de sistemas basados en inteligencia artificial ha superado la barrera del campo

académico debido a sus potencialidades en la gestión pública. Su investigación ofreció una visión panorámica del impacto de la IA en el campo de la gestión y administración pública, abordando logros y controversias significativas. El estudio identificó oportunidades y desafíos críticos de aplicación de la IA en el sector público, proporcionando un marco conceptual para la evaluación de impactos.

Estudios adicionales han explorado la aplicación de inteligencia artificial en sectores específicos como salud pública, gestión de recursos, análisis predictivo y automatización de procesos administrativos (Castaño Castaño, 2025). Estas investigaciones documentaron beneficios tangibles en términos de eficiencia operativa, mejora en la toma de decisiones y optimización de recursos, pero también identificaron desafíos persistentes relacionados con la interpretabilidad de algoritmos, gestión de la privacidad y mantenimiento de la supervisión humana. Los estudios revelan que la incorporación de IA en administración pública constituye una transformación relevante en el ámbito gubernamental contemporáneo, requiriendo marcos de implementación responsable basados en diagnóstico, diseño, piloto, despliegue y evaluación continua.

2.4.Casos de Implementación en América Latina

La implementación práctica de sistemas de inteligencia artificial en el sector público colombiano y latinoamericano ha generado insumos empíricos relevantes para identificar oportunidades, limitaciones y aprendizajes derivados de su aplicación en contextos institucionales concretos. Estos casos ofrecen una visión comparativa sobre los avances alcanzados y los desafíos pendientes en la integración de estas tecnologías en la gestión pública.

2.4.1. Casos de Implementación en Colombia

Ruiz Sánchez (2024) examinó la incorporación de inteligencia artificial en las Altas Cortes de Colombia como uno de los ejemplos más representativos de adopción en el ámbito judicial. Los proyectos adelantados en la Corte Constitucional y el Consejo de Estado, desarrollados con el apoyo de juristas de la Universidad del Rosario y del laboratorio IALAB de la Universidad de Buenos Aires, demostraron que la inteligencia artificial no solo permite mejorar la eficiencia en la elaboración de decisiones judiciales, sino que también influye en la cultura jurídica y en los procesos organizacionales. Los resultados obtenidos confirman avances en la eficiencia judicial y fortalecen principios asociados con el acceso a la justicia, la tutela efectiva y la protección de derechos fundamentales.

Desde una perspectiva de política pública, Gutiérrez Peña (2024) realizó un balance de las iniciativas de inteligencia artificial y transformación digital implementadas entre 2019 y 2024. El análisis evidenció que Colombia se consolidó como pionera regional en el diseño de marcos regulatorios, en especial con la adopción del CONPES 3975 de 2019 (Consejo Nacional de Política Económica y Social,

2019) y su actualización a través del CONPES 4144 (2024). No obstante, la evaluación señala falta de continuidad en el desarrollo de políticas y ausencia de mecanismos de seguimiento que ponen en riesgo la sostenibilidad del liderazgo alcanzado. En consecuencia, el autor propone una agenda de evaluación que subsane estas debilidades y garantice una implementación más sistemática de las iniciativas en el sector público.

En otra línea, Martínez González et al. (2024) analizaron la percepción de la inteligencia artificial en la lucha contra la corrupción en el Estado colombiano. Los resultados muestran que la totalidad de los encuestados considera indispensable ampliar el uso de herramientas tecnológicas para enfrentar este problema, destacando su potencial para mejorar la transparencia y la eficiencia administrativa. Sin embargo, la investigación identificó limitaciones significativas en capacidades organizacionales, con bajos puntajes en liderazgo directivo y en gestión de inteligencia artificial, lo que dificulta una aplicación efectiva de estas herramientas.

En el plano normativo y ético, el Instituto Colombiano de Normas Técnicas y Certificación (2023) formuló un marco orientador para el uso responsable de la inteligencia artificial en entidades públicas. Este documento establece principios de equidad, transparencia, responsabilidad y participación ciudadana, y define directrices para asegurar que la implementación de sistemas de inteligencia artificial en servicios gubernamentales se realice bajo criterios éticos claros.

2.4.2. Casos de Implementación en LATAM

En el ámbito regional, la Carta Iberoamericana de Inteligencia Artificial en la Administración Pública constituye un esfuerzo colectivo para establecer lineamientos comunes sobre la implementación de estas tecnologías en el sector público (Revista CLAD, 2024). El análisis de cuestionarios aplicados a responsables nacionales de servicio civil y modernización administrativa mostró una fuerte heterogeneidad en el grado de madurez y adopción entre los países. En este contexto, Brasil, Chile, México, Perú y Colombia destacan como referentes regionales por el uso de esquemas de colaboración público-privada que han permitido cerrar brechas de inversión y mejorar la eficiencia de los servicios estatales.

Un análisis comparativo realizado por González Méndez y Vásquez López (2024) sobre los modelos de gobernanza de inteligencia artificial en América del Norte identificó diferencias sustanciales entre países. Estados Unidos privilegia un enfoque centrado en la seguridad nacional con predominio de iniciativas estatales, Canadá promueve la colaboración entre gobierno, sector privado y academia con un enfoque humano y orientado al crecimiento económico, mientras que México presenta rezagos en políticas nacionales, aunque se observa un rol activo de la sociedad civil en la discusión y formulación de lineamientos en esta materia.

De manera específica, Rivera Hernández (2024) estudió la aplicación de inteligencia artificial en el sector público mexicano, con énfasis en la simplificación de procesos administrativos y en el análisis de bases de datos tributarias. El Servicio de Administración Tributaria incorporó inteligencia artificial en el Plan Maestro 2024 como herramienta para segmentar contribuyentes por nivel de riesgo, identificar redes de evasión y contrabando, y auditar operaciones de empresas fachada (Fernández Torres, 2025). Estos usos consolidaron la inteligencia artificial como un recurso estratégico en la fiscalización tributaria.

En el campo de la salud, Restrepo Silva y González Vargas (2024) documentaron experiencias de asociaciones público-privadas en Brasil, Chile, México, Perú y Colombia orientadas a la incorporación de inteligencia artificial en servicios sanitarios. Los resultados muestran beneficios en eficiencia y calidad de la atención, aunque persisten limitaciones relacionadas con capacidad técnica, marcos regulatorios y resistencia institucional al cambio.

En el contexto latinoamericano, Zamora Pérez (2025) analizó la implementación de inteligencia artificial en la administración pública, documentando casos en Estonia, Brasil, Chile y Ecuador. La investigación sostiene que la integración de IA representa una transformación estructural en la gestión pública, que requiere procesos de implementación gradual basados en diagnóstico, diseño, pruebas piloto, despliegue y evaluación permanente.

Los casos latinoamericanos dejan ver logros en eficiencia operativa, optimización de recursos y mejora en la toma de decisiones, pero también reflejan desafíos vinculados con la explicabilidad de los algoritmos, la protección de datos y la necesidad de mantener supervisión humana en el proceso de toma de decisiones. Estas experiencias resaltan la importancia de desarrollar marcos regulatorios más específicos, fortalecer capacidades técnicas y diseñar estrategias de gestión del cambio organizacional que permitan potenciar el valor transformador de la inteligencia artificial en el sector público.

2.5. Metodología de Revisión Sistemática

La selección de literatura para esta revisión siguió una metodología sistemática basada en criterios explícitos de inclusión y exclusión, garantizando la relevancia, calidad y actualidad de las fuentes analizadas. La metodología adoptada se fundamenta en los lineamientos establecidos por Kitchenham & Charters (2007) para revisiones sistemáticas, adaptados al contexto específico de la gobernanza de inteligencia artificial en el sector público.

Tabla 2. Protocolo de revisión sobre gobernanza de la IA en el sector público

Etapa	Puntos Clave
Matriz de Extracción	5 ejes: Ética/Derechos, Riesgo/Controles, Transparencia, Instrumentos, Evidencia Pública.
Estrategia de Búsqueda	Cadenas booleanas (ES/EN), Periodo 2020–2025, Acceso Abierto, Bases Académicas/Repositorios.
Fuentes	Bases académicas, Repositorios (OCDE, UNESCO, NIST, etc.), Revistas y Portales de Acceso Abierto.
Criterios de Inclusión	Foco en **Gobernanza/Ética/Política de IA** en el **Sector Público**, Instrumentos Aplicables, Publicaciones **2020-2025**, Acceso Abierto.
Criterios de Exclusión	Tesis/Preprints sin metodología, Artículos técnicos (no política pública), Duplicados, Sin trazabilidad.
Proceso PRISMA	**247** identificados -> **180** cribados -> **86** evaluados -> **59** incluidos.
Evaluación de Calidad	Robustez metodológica, Pertinencia temática (IA en sector público), Credibilidad, Riesgo de sesgo.

Fuente: Elaboración Propia

El proceso de selección adoptó un flujo PRISMA con cribado por título y resumen, lectura a texto completo y decisión final, registrando los siguientes totales: 247 registros identificados, 67 duplicados eliminados, 180 registros cribados por título/resumen, 94 excluidos por irrelevancia temática o metodológica, 86 evaluados a texto completo e inclusión final de 59 estudios (Page et al., 2021).

La aplicación sistemática de estos criterios resultó en la selección de 104 fuentes web primarias, complementadas con documentos oficiales de marcos regulatorios y estándares internacionales. Esta metodología garantiza que la revisión bibliográfica proporcione una base sólida y actualizada para el desarrollo del framework de gobernanza propuesto, reflejando tanto los avances teóricos como las lecciones prácticas de implementación en contextos gubernamentales reales, con particular énfasis en la aplicabilidad al contexto institucional del Distrito Capital de Bogotá.

2.6. Análisis Comparativo y Brechas de Implementación

Los marcos de referencia más influyentes convergen en principios de derechos humanos, proporcionalidad de riesgo, transparencia y rendición de cuentas, pero divergen en su preparación para contextos con capacidades limitadas y arreglos institucionales fragmentados a nivel subnacional (OECD, 2024). Esta divergencia se manifiesta en el grado de prescriptividad, en la carga de cumplimiento, en la disponibilidad de instrumentos listos para usar y en la existencia de métricas

obligatorias versus recomendadas, lo cual es crítico para municipalidades y entidades con equipos generalistas (NIST, 2023). Este análisis comparativo se detalla en el Anexo A.

2.6.1. Limitaciones en la implementación

La mayor parte de estos marcos se concibió para contextos nacionales o supranacionales con ecosistemas tecnológicos y arreglos de supervisión robustos, lo que genera tensiones al trasladarlos a gobiernos locales con escalas organizacionales y presupuestales reducidas (OECD, 2024). Sin una metodología de adaptación que traduzca obligaciones y procesos a capacidades reales, es probable que se produzcan fricciones, cumplimiento formal poco sustantivo o abandono de iniciativas tras pilotos iniciales (ALSUR, 2024).

Un eje crítico es la transparencia algorítmica aplicable, donde la experiencia chilena sugiere que catálogos de algoritmos, estándares de divulgación y participación multiactor son componentes habilitantes para pasar de principios a prácticas, con instrumentos que permiten escrutinio público y aprendizaje institucional (OECD OPSI, 2023). Del mismo modo, las series comparativas de la OCDE en la región y diagnósticos específicos muestran que las brechas de habilidades, gobernanza de datos y evaluación de impacto dificultan la escalabilidad y sostenibilidad de proyectos de IA en el sector público (OECD, 2024).

Las evaluaciones de impacto algorítmico ofrecen un camino intermedio entre marcos prescriptivos y prácticas voluntarias, al introducir listas de verificación, documentación de propósito, análisis de riesgos y vías de reclamación, especialmente útiles en dominios sensibles como salud o bienestar social (Ada Lovelace Institute, 2022). En América Latina, los análisis recientes sobre automatización de prestaciones sociales evidencian el valor de integrar medición de resultados, supervisión independiente y mecanismos de apelación, con énfasis en el sesgo, la explicabilidad y los impactos distributivos (Internet Policy Review, 2024).

2.6.2. Evidencia y tendencias regionales relevantes

Los mapeos regulatorios y comparativos latinoamericanos identifican rutas diversas que van desde principios generales y hojas de ruta hasta proyectos de ley con registros de sistemas, evaluaciones de impacto y autoridades dedicadas, con desafíos de coordinación interinstitucional y capacidades técnicas (Access Now, 2023). En paralelo, iniciativas multilaterales han propuesto marcos de referencia interamericanos para gobernanza de datos e IA que buscan armonizar principios y facilitar cooperación regional, aspecto clave para compartir costos de capacidad y auditoría (OEA, 2025).

A nivel nacional y subnacional, compromisos de gobierno abierto y programas formativos para el servicio civil han servido como palancas para construir capacidades, generar inventarios de algoritmos

y lanzar pilotos de transparencia y auditoría adaptados a contextos institucionales concretos (UNESCO, 2025). En Colombia, agendas recientes proponen construir un modelo multiactor de gobernanza de IA articulado con la política nacional, con resultados esperados de institucionalización y clarificación de roles en niveles estratégicos, tácticos y operativos (OGP Colombia, 2025).

Los índices de gobierno digital y estudios de línea base para América Latina y el Caribe muestran heterogeneidad marcada en madurez, lo que aconseja escalonar herramientas y métricas, priorizando primero capacidades habilitantes de datos, talento y gobernanza de proyectos, y luego mecanismos avanzados de certificación o auditoría (OECD & IDB, 2024). En síntesis, la evidencia sugiere privilegiar "proporcionalidad operacional": marcos y herramientas deben amoldarse a la capacidad instalada y al ciclo presupuestal real de las entidades, especialmente en municipios y ciudades intermedias (OECD, 2024).

2.7.Contribución Esperada y Originalidad

A partir del análisis sistemático realizado que evidencia claramente una brecha en cuanto a la desconexión del marco normativo colombiano, haciendo referencia particularmente al documento CONPES 4144 (2024) en contraste con los instrumentos operativos requeridos para su implementación efectiva en entidades distritales de Bogotá (Departamento Nacional de Planeación, 2025). Tres limitaciones convergentes configuran esta brecha. La primera es la falta de adaptación contextual, ya que los marcos existentes no reconocen ni se ajustan adecuadamente a las particularidades institucionales, técnicas y políticas del Distrito Capital. Factores estructurales como capacidades heterogéneas entre entidades, la dependencia de proveedores tecnológicos y la discontinuidad política no son tangibles, lo que compromete la aplicabilidad práctica de los frameworks actuales.

La segunda limitación corresponde a la fragmentación implementativa debido a que no se observa una integración sistemática entre las cinco dimensiones críticas de la gobernanza de IA: principios éticos, requisitos normativos, controles técnicos, herramientas operativas y métricas de evaluación. Esta fragmentación genera esfuerzos dispersos que no logran capturar la complejidad ni garantizar la coherencia de una gobernanza responsable.

La tercera limitación es el déficit de validación empírica. Los marcos disponibles carecen de evidencia robusta sobre su efectividad en contextos subnacionales reales, lo que restringe su aplicabilidad práctica en entidades con las características propias del sector público distrital lo que hacen que su aplicabilidad sea operativamente inviable.

Es clave destacar que la brecha es aplicable dentro del contexto del Distrito de Bogotá y que está articulada a la normativa (CONPES, 2024) y que su orientación implementativa está mas enfocada a

herramientas operativas que en principios más abstractos. Debido a la reciente tendencia de adopción de sistemas de IA en servicios públicos se justifica una necesidad inmediata de establecer un gobierno focalizado que atienda de manera adecuada tanto las nuevas capacidades como las existentes.

2.7.1. Propuesta del Framework

La investigación propuesta buscar responder de manera directa la brecha identificada mediante el desarrollo de un framework de gobernanza estructurado en cinco capas interoperables: principios éticos, requisitos normativos, controles técnicos, herramientas operativas y métricas de evaluación. Esta arquitectura asegura tanto la trazabilidad de la normativa (CONPES, 2024) como a los marcos internacionales relevantes, por lo que a su vez habilita una implementación flexible y adaptada a capacidades reales en el ámbito de entidades distritales, con lo que se supera la rigidez de los enfoques universalistas.

La originalidad de esta propuesta radica en su capacidad para operar de manera simultánea como instrumento de cumplimiento normativo alineado a los mandatos (CONPES, 2024), paralelamente como herramienta de gestión práctica que no requiera de capacidades técnicas excepcionales, y también como un sistema adaptativo que se ajusta dinámicamente a la heterogeneidad institucional del contexto bogotano.

Como parte del alcance de este trabajo se desarrollará un sistema integrado de herramientas operativas diseñado exclusivamente para el entorno técnico, político e institucional del Distrito Capital. Dentro sistema se incluirán herramientas como matrices de evaluación de riesgos contextualizadas para los procesos gubernamentales, plantillas de evaluación de impacto algorítmico adaptadas a recursos institucionales disponibles, protocolos de auditoría compatibles con capacidades técnicas existentes, matrices de roles y responsabilidades ajustados a estructuras organizacionales distritales, y métricas de cumplimiento normativo verificables con herramientas disponibles.

El núcleo innovador de la propuesta viene con la integración sistémica, la contextualización específica y la validación empírica, ya que son elementos que representan una contribución original al campo de gobernanza de IA en contextos subnacionales.

Por último, se diseñará una metodología de validación multinivel, que contará con la participación de funcionarios distritales responsables de implementación, consultores expertos en gobernanza de IA, y los mismos ciudadanos usuarios de los servicios públicos. Esta triangulación garantiza que el framework responda tanto a exigencias técnicas, como operativas y democráticas, y que asegura su legitimidad y viabilidad práctica.

2.8.Transición al Marco Metodológico

Los hallazgos derivados del estado del arte constituyen insumos directos para el diseño metodológico de la investigación, al establecer requerimientos específicos que deben ser abordados por el framework propuesto. En primer lugar, la heterogeneidad institucional observada en los estudios revisados demanda la adopción de metodologías participativas que integren las perspectivas de entidades con capacidades diferentes, evitando así soluciones estandarizadas que desatiendan dichas diferencias (Filgueiras, 2023a). En segundo término, el análisis de Mendonça et al. (2023) sugiere que las brechas entre la gobernanza tradicional y la automatización requieren enfoques que medien entre la rigidez de la lógica algorítmica y la necesidad de validación democrática. De este modo, las propuestas teóricas sobre institucionalismo digital no solo deben buscar eficiencia operativa, sino garantizar que los algoritmos actúen como reglas transparentes y legítimas dentro del tejido social. Asimismo, la convergencia normativa detectada habilita el uso de metodologías de mapeo que garanticen trazabilidad entre principios internacionales, políticas nacionales y herramientas locales, tal como lo recomienda la Unesco (2021). Finalmente, la evidencia empírica sobre casos exitosos permite incorporar metodologías comparativas orientadas a identificar factores críticos de éxito. Siguiendo a Evans (1995), este análisis no solo busca la replicabilidad técnica, sino comprender cómo la capacidad institucional y el enraizamiento de las políticas en el tejido social de Bogotá son determinantes para la transformación efectiva del entorno local.

Resulta fundamental señalar que el proceso de validación metodológica deberá incluir componentes explícitos de transferencia de conocimiento y gestión del cambio institucional. La efectividad del framework, en este sentido, dependerá no solo de su calidad técnica, sino también de su apropiación por parte de las entidades involucradas, un aspecto que la literatura técnica tiende a subestimar sistemáticamente (Heeks, 2002). Esta transición metodológica será desarrollada de manera integral en el Capítulo 3, donde se presentará el diseño específico que operacionaliza los requerimientos identificados en el análisis del estado del arte, proporcionando así la base empírica y conceptual para el desarrollo del framework propuesto.

3. Objetivos y Metodología de Trabajo

Este capítulo establece los objetivos y la metodología que guiarán el desarrollo del Framework de Gobernanza de IA para el sector público del Distrito Capital de Bogotá. Sirve como puente esencial entre el diagnóstico del problema y el análisis del estado del arte, presentados en los capítulos anteriores, y el diseño práctico de la contribución de esta investigación.

3.1. Objetivo Principal

[OP] Diseñar, desarrollar y validar un Framework de Gobernanza de IA integral y práctico, junto con su caja de herramientas asociada, para permitir a las entidades públicas del Distrito Capital de Bogotá adoptar e implementar sistemas de inteligencia artificial de manera ética, segura, legal y efectiva. Este objetivo responde directamente a la brecha de implementación identificada entre el marco normativo nacional (CONPES, 2024) y las capacidades operativas reales a nivel subnacional (Departamento Nacional de Planeación, 2025), integrando y adaptando críticamente los principios de los marcos internacionales más relevantes (UE, OCDE, UNESCO) al contexto institucional bogotano.

3.2. Objetivos Específicos

Para alcanzar el objetivo principal, se han definido los siguientes objetivos específicos:

[OE1] Realizar un diagnóstico integral del ecosistema local de gobernanza de IA en el Distrito Capital, mediante la comparación de sus capacidades institucionales, brechas operativas y marcos normativos vigentes con estándares internacionales, para establecer una línea base de diseño.

[OE2] Diseñar la arquitectura integral del Framework de Gobernanza de IA, estructurado en cinco capas interoperables (Principios y Políticas, Modelo de Gobierno, Fases del Ciclo de Vida, Controles Clave, y Métricas/KPIs), asegurando su trazabilidad con la normativa nacional y su adaptabilidad al contexto institucional del Distrito.

[OE3] Desarrollar la caja de herramientas operativa (toolkit) del framework, que incluye instrumentos estandarizados como el AI Use-Case Canvas, la Matriz de Riesgos de IA, la Plantilla ARA/DPIA, Data Sheets, el Checklist de Evaluación de Proveedores, la Model Card y la Guía de Uso Interno de IA Generativa, priorizando la usabilidad y adaptabilidad a las capacidades técnicas heterogéneas de las entidades distritales.

[OE4] Validar la efectividad y usabilidad del framework y su toolkit mediante la ejecución de una simulación sobre un servicio público distrital, midiendo su impacto en el cumplimiento normativo, la eficiencia del servicio y la gestión de riesgos.

[OE5] Consolidar los hallazgos de la validación, refinar los artefactos desarrollados y elaborar una guía de implementación escalonada con un modelo de madurez institucional, para facilitar la adopción progresiva del framework en el sector público distrital.

3.3. Metodología de Trabajo

La metodología se articula en cuatro fases secuenciales, alineadas con los objetivos específicos y basadas en el paradigma de Design Science Research (DSR), el cual es idóneo para la creación y validación de artefactos que resuelven problemas organizacionales identificados (Hevner et al., 2004).

3.3.1. Fase 1: Análisis Contextual y de Requisitos

Descripción:

Esta fase constituye la base diagnóstica del proyecto y tiene como propósito consolidar un entendimiento profundo del contexto institucional, normativo y operativo del Distrito Capital de Bogotá en materia de inteligencia artificial. A partir de un análisis sistemático, se identifican las brechas entre el marco normativo nacional (CONPES 4144, 2024) y las capacidades reales de las entidades distritales, y se extraen los requisitos funcionales y no funcionales que deben satisfacerse mediante el diseño del framework. Esta fase logra el objetivo específico OE1 y proporciona los insumos necesarios para el diseño contextualizado de la arquitectura y el toolkit.

Actividades:

La fase se desarrolló mediante un proceso estructurado en cuatro actividades principales. Primero, se realizó un análisis del contexto distrital y un mapeo de capacidades institucionales, que incluyó la revisión documental de planes estratégicos, informes de gestión y estructuras organizacionales de las entidades del Distrito Capital, así como la identificación de iniciativas previas o en curso relacionadas con IA, transformación digital y gobierno de datos, y la caracterización de la heterogeneidad en capacidades técnicas, recursos humanos e infraestructura tecnológica entre entidades. Segundo, se ejecutó una evaluación de brechas normativas y operativas, mediante el análisis comparativo entre los mandatos del CONPES 4144 (2024) y los instrumentos operativos disponibles en las entidades distritales, la identificación de limitaciones estructurales como la dependencia de proveedores externos, la discontinuidad política y la fragmentación implementativa, y el diagnóstico de la madurez en gobernanza de datos, privacidad y gestión de riesgos tecnológicos. Tercero, se llevó a cabo una revisión sistemática de marcos internacionales y buenas prácticas, que comprendió el análisis comparativo estructurado de marcos de referencia como el AI Act de la UE, el NIST AI RMF, los Principios de la OCDE, la Recomendación de la UNESCO y los estándares ISO/IEC 23894 e ISO/IEC

42001, la identificación de componentes transferibles y adaptables al contexto subnacional bogotano, y la extracción de principios, controles y métricas relevantes para la gobernanza de IA en el sector público. Cuarto, se priorizaron los requisitos para el diseño del framework, a través de la consolidación de una lista de requisitos funcionales y no funcionales, su clasificación según criticidad y viabilidad de implementación, y la definición de criterios de éxito para el framework y su toolkit asociado.

Instrumentos y Técnicas:

Para el desarrollo de estas actividades se emplearon matrices de análisis comparativo de marcos normativos y plantillas de extracción de requisitos basadas en estándares internacionales. La principal técnica fue la revisión documental exhaustiva de políticas distritales, informes de auditoría, planes de transformación digital y literatura académica relevante. El análisis y la sistematización de los hallazgos se realizaron mediante síntesis manual y tablas comparativas.

Validación:

Los resultados de esta fase se validaron mediante la triangulación metodológica entre la evidencia documental recopilada de fuentes oficiales, la consistencia con los hallazgos del estado del arte presentado en el Capítulo 2, y la revisión crítica por parte del equipo investigador y el director del trabajo. Este proceso aseguró la robustez y pertinencia del diagnóstico, cuyo producto final es un Informe de Diagnóstico y Requisitos que sirve como línea base para el diseño del framework y que fundamenta las decisiones de arquitectura y desarrollo del toolkit en la Fase 2.

3.3.2. Fase 2: Diseño del Framework y Desarrollo del Toolkit

Descripción:

Esta fase materializa la propuesta al traducir los principios, requisitos y hallazgos diagnósticos de la Fase 1 en una arquitectura de gobernanza operable y un conjunto de herramientas prácticas. Su propósito es construir los artefactos centrales de la investigación: un framework estructuralmente sólido y un toolkit listo para su uso institucional. Esta fase logra de manera integral los objetivos específicos OE2 y OE3, asegurando que el diseño no solo sea teóricamente robusto sino también aplicable al contexto distrital.

Actividades:

La fase se articuló en dos líneas de trabajo paralelas e interconectadas. Primero, se procedió al diseño de la arquitectura integral del Framework de Gobernanza de IA, lo que implicó la definición y especificación detallada de sus cinco capas interoperables. Para la Capa 1 (Carta de IA Responsable y Políticas) se redactaron los principios fundamentales y las políticas de gobierno de datos, gestión de riesgos y compras, alineándolas con la Constitución, la Ley 1581 de 2012 y el CONPES 4144. En la Capa 2 (Modelo de Gobierno) se diseñó la estructura organizacional, definiendo las modalidades de Comité

de IA (distrital o por entidad), los roles operativos a nivel de caso de uso y la matriz de responsabilidades RACI. La Capa 3 (Fases del Ciclo de Vida) se estructuró en nueve etapas secuenciales, desde el *intake* hasta el retiro, incorporando puntos de control (gates) y entregables obligatorios. Para la Capa 4 (Controles Clave) se definieron seis dimensiones de controles técnicos y organizacionales: ética y derechos, privacidad, seguridad, transparencia, atención al ciudadano y gestión de terceros. Finalmente, la Capa 5 (Métricas y KPIs) se diseñó integrando un sistema de medición basado en COBIT 2019, con métricas de cumplimiento, gestión de riesgos, calidad de datos, desempeño técnico, impacto en el servicio y madurez institucional. En segundo lugar, se desarrolló la caja de herramientas operativa (toolkit), creando y prototipando siete instrumentos estandarizados. Se diseñó el AI Use-Case Canvas para la fase de *intake*, la Matriz de Riesgos de IA con evaluación probabilidad-impacto, la Plantilla integrada ARA/DPIA, el Data Sheet para documentación de conjuntos de datos, el Checklist de Evaluación de Proveedores en tres versiones según riesgo, la Model Card para transparencia técnica y la Guía de Uso Interno de IA Generativa. Cada herramienta se desarrolló con manuales de usuario, formatos diligenciables y se aseguró su integración coherente con las fases del ciclo de vida del framework.

Instrumentos y Técnicas:

Para el diseño arquitectónico se utilizaron herramientas de diagramación como PlantUML, Draw.io y software de modelado de procesos. El desarrollo del toolkit se realizó mediante suites ofimáticas avanzadas y formatos HTML + JavaScript para crear plantillas interactivas con validaciones y cálculos automáticos. Todo el proceso se apoyó en la referencia constante a estándares internacionales como ISO/IEC 23894, ISO/IEC 42001 y el NIST AI RMF para garantizar el rigor técnico y la trazabilidad normativa.

Validación:

La validación en esta fase fue de consistencia interna y coherencia conceptual. Se realizaron revisiones iterativas para asegurar que cada componente del framework y cada herramienta del toolkit estuvieran alineados con los requisitos identificados en la Fase 1 y fueran trazables a los marcos de referencia nacionales e internacionales. Se verificó la integridad sistémica del diseño, confirmando que las cinco capas fueran interoperables y que el toolkit ofreciera soporte práctico a todo el ciclo de vida de gobernanza.

3.3.3. Fase 3: Validación Experimental y Refinamiento

Descripción:

Esta fase busca obtener evidencia empírica sobre la aplicabilidad, utilidad y efectividad de los artefactos desarrollados (framework y toolkit) en un contexto simulado que refleje la realidad

operativa del Distrito Capital. Su propósito es probar el funcionamiento integral de la propuesta, identificar ajustes necesarios y medir su impacto preliminar en indicadores clave. Esta fase logra el objetivo específico OE4, superando el déficit de validación empírica identificado en el estado del arte y proporcionando una base sólida para la refinación de los artefactos.

Actividades:

La fase se centró en la ejecución de un proceso de validación experimental mediante simulación. Inicialmente, se seleccionó un caso de uso representativo de un servicio público distrital, específicamente un chatbot de atención ciudadana y un sistema de priorización de trámites, por su relevancia y potencial para evidenciar los mecanismos de gobernanza. Sobre este caso, se aplicó de manera integral el proceso de ciclo de vida completo definido en el framework, utilizando todas las herramientas del toolkit. Esto implicó la diligenciación del AI Use-Case Canvas, la clasificación de riesgo, la realización del ARA/DPIA, la gestión de datos con Data Sheets, la evaluación de proveedores con el checklist, la documentación del modelo con la Model Card y la aplicación de la guía de IA generativa donde correspondía. Durante la simulación, se midieron resultados a través de KPIs predefinidos, incluyendo tiempos de respuesta, tasa de cumplimiento de controles del DPIA, nivel de satisfacción de usuario simulado y número de incidentes o desviaciones identificadas. Para generar los datos de entrada y los escenarios de prueba de manera realista y controlada, se utilizó Inteligencia Artificial Generativa para crear perfiles de ciudadanos, narrativas de interacción, conjuntos de datos sintéticos y respuestas simuladas de los sistemas. El proceso fue documentado en detalle, generando los formatos diligenciados que se presentan en el Anexo E. Finalmente, se realizó un análisis cualitativo de los comentarios y hallazgos emergentes durante la simulación, extrayendo lecciones aprendidas sobre usabilidad, claridad de las herramientas y puntos de fricción en el proceso.

Instrumentos y Técnicas:

Para la simulación se utilizaron plataformas de prototipado y entornos de desarrollo como Visual Studio junto con herramientas de co-creación asistida por IA (Copilot Studio) para emular el comportamiento de los sistemas. La generación de los datos y escenarios de prueba se apoyó en modelos de lenguaje de IA Generativa para crear contenido sintético realista y contextualizado. El análisis de los resultados combinó métodos cuantitativos (cálculo de KPIs a partir de los datos simulados) y cualitativos (análisis temático del feedback generado durante el proceso simulado).

Validación:

La validación en esta fase se fundamentó en un doble criterio. Primero, en el consenso alcanzado entre los investigadores y los roles simulados (sponsor, técnico, DPO, etc.) sobre la usabilidad, claridad y utilidad percibida de los artefactos. Segundo, en la mejora cuantificable y observable de los KPIs

medidos durante la simulación, en comparación con una línea base teórica o con procesos no estructurados. La combinación de evidencia cuantitativa y cualitativa permitió establecer una validación práctica robusta, confirmando que el framework y el toolkit no solo son conceptualmente sólidos, sino también operativamente viables y beneficiosos.

3.3.4. Fase 4: Consolidación y Elaboración de la Guía de Implementación

Descripción:

Esta fase final de síntesis tiene como propósito incorporar el *feedback* y las lecciones aprendidas de la validación experimental, refinar los artefactos desarrollados y empaquetar la contribución completa de la investigación en un formato listo para su transferencia y aplicación práctica por parte de las entidades del Distrito. Su objetivo es asegurar que el framework no sea solo un producto académico, sino una solución implementable que responda a la problemática identificada. Esta fase logra el objetivo específico OE5, cerrando el ciclo de la metodología DSR con un artefacto completo, validado y acompañado de una hoja de ruta para su adopción.

Actividades:

La fase comprendió tres actividades principales de consolidación. En primer lugar, se refinó el framework y el toolkit incorporando de manera sistemática las lecciones aprendidas y los ajustes identificados durante la Fase 3 de validación. Esto incluyó la clarificación de instrucciones, el ajuste de umbrales en algunas métricas, la simplificación de flujos donde se identificó complejidad innecesaria y el fortalecimiento de las referencias cruzadas entre herramientas. En segundo lugar, se elaboró una Guía de Implementación del Ciclo de Vida de Gobernanza de IA, la cual fue desarrollada en detalle y estructurada como el Anexo D del presente trabajo. Esta guía integral proporciona instrucciones paso a paso, responsabilidades, tiempos y formatos asociados para cada una de las nueve fases del ciclo de vida, facilitando su aplicación operativa por parte de los equipos distritales. Además, se diseñó una guía de implementación escalonada a nivel estratégico, que incluye un plan de trabajo a 12 meses y un modelo de madurez institucional con cuatro niveles (Incipiente, Informal, Documentado y Optimizado) que permite a las entidades autoevaluarse y trazar su ruta de avance. En tercer lugar, se definieron las especificaciones técnicas y funcionales para un tablero de control (dashboard) de monitoreo, cuyas definiciones conceptuales, métricas y estructura se documentan en el Anexo C. Este anexo establece la base para el desarrollo futuro de la herramienta de visualización. Adicionalmente, se consolidó toda la documentación generada—incluyendo el framework, el toolkit, los manuales, los anexos y las guías—en la estructura final del presente trabajo, asegurando coherencia narrativa y exhaustividad.

Instrumentos y Técnicas:

Para esta fase se emplearon herramientas de documentación y gestión de contenidos para la redacción unificada y estructuración de la guía de implementación (Anexo D) y las definiciones del dashboard (Anexo C). La técnica central fue la síntesis documental y el diseño instruccional para transformar los artefactos validados en materiales de transferencia práctica. La gestión del conocimiento se apoyó en repositorios estructurados para versionar todos los artefactos y documentos asociados, asegurando la trazabilidad entre los diferentes componentes del framework y sus anexos.

Validación:

La calidad, completitud y utilidad de los entregables finales de esta fase se validó mediante una revisión exhaustiva final. Esta revisión aseguró que la solución integral —framework, toolkit, guía de implementación (Anexo D), modelos de madurez y definiciones del dashboard (Anexo C)— constituya una respuesta coherente, práctica y lista para su implementación que aborde de manera integral la problemática y la brecha de investigación identificadas al inicio del trabajo. La validación confirmó que los artefactos refinados están alineados con los objetivos de la investigación, son aptos para su transferencia al contexto del sector público distrital y cuentan con la documentación de soporte necesaria para su adopción exitosa.

4. Desarrollo de la Propuesta y Análisis de Resultados

Este capítulo presenta la propuesta de Framework de Gobernanza de IA desarrollada para las entidades públicas del Distrito Capital de Bogotá, junto con los resultados de su proceso de validación. Se estructura en dos grandes secciones: la primera expone en detalle la arquitectura del framework y su caja de herramientas operativa (toolkit); la segunda presenta el análisis de los resultados obtenidos mediante la validación con expertos y la ejecución de pilotos controlados.

4.1. Arquitectura del Framework de Gobernanza de IA

El framework propuesto constituye un sistema integrado de gobernanza estructurado en cinco capas interoperables que traducen los principios del marco nacional (CONPES, 2024) y los marcos internacionales en instrumentos operativos adaptados al contexto institucional del Distrito Capital de Bogotá. Esta arquitectura responde directamente a las tres limitaciones identificadas en el estado del arte: falta de adaptación contextual, fragmentación implementativa y déficit de validación empírica.

En el Anexo B se ilustra la arquitectura general del framework, mostrando las secciones de cada una de sus cinco capas y su alineación con la normativa nacional (CONPES, 2024).

4.1.1. Capa 1: Carta de IA Responsable y Política de Uso de IA

La Carta de IA Responsable constituye el documento declarativo que establece el compromiso institucional del Distrito con una implementación ética de sistemas de inteligencia artificial. Este documento se fundamenta en estándares internacionales de la UNESCO (2021) y OCDE (2024), adaptados específicamente al marco normativo colombiano mediante su articulación con la Constitución Política, la Ley de protección de datos personal (Ley 1581, 2012) y la normativa nacional (CONPES, 2024).

Este marco normativo se complementa con la Ley 2279 (2022) de Transformación Digital, que establece las bases para la modernización del Estado mediante tecnologías emergentes, habilitando jurídicamente la adopción de sistemas automatizados en servicios públicos. Adicionalmente, el framework incorpora flexibilidad para integrar las disposiciones de la futura Ley de Inteligencia Artificial actualmente en trámite legislativo, asegurando adaptabilidad normativa prospectiva.

Los principios fundamentales establecidos son:

Respeto por los derechos humanos y dignidad. Todo sistema de IA implementado en servicios públicos distritales debe respetar los derechos fundamentales consagrados en la Constitución, con especial énfasis en el derecho a la igualdad (Art. 13), el habeas data o protección de datos personales (Art. 15), el debido proceso (Art. 29) y el acceso a servicios públicos (Art. 365). Este principio se

materializa mediante evaluaciones obligatorias de impacto antes del despliegue de sistemas que afecten decisiones sobre ciudadanos.

Transparencia y explicabilidad. Las entidades distritales deberán implementar mecanismos de comunicación proactiva que informen a la ciudadanía cuando estén interactuando con sistemas de inteligencia artificial. Esta comunicación deberá incluir explicaciones claras y comprensibles sobre el propósito, funcionamiento y criterios de decisión de estos sistemas.

Para sistemas de alto riesgo, se requiere documentación técnica accesible mediante Model Cards y Data Sheets, junto con mecanismos de explicación adaptados a cada perfil de usuario: ciudadanos recibirán información simplificada, funcionarios contarán con datos técnicos para supervisión, y auditores dispondrán de documentación completa para verificación del cumplimiento normativo.

Rendición de cuentas y supervisión humana. Se establece el principio de que los sistemas de IA complementarán pero no reemplazarán el criterio humano en decisiones que afecten derechos fundamentales o servicios esenciales. Cada sistema deberá incorporar mecanismos de supervisión humana efectiva que permitan la intervención, modificación o revocación de decisiones algorítmicas cuando sea necesario.

Para garantizar esta supervisión, se designarán responsables específicos para cada implementación de IA, estableciendo una cadena clara de responsabilidades que incluye al sponsor del área usuaria, el responsable técnico y el Comité de IA correspondiente. Esta estructura asegura la debida rendición de cuentas y mantiene el control institucional sobre los sistemas automatizados.

Equidad y no discriminación. Los sistemas de IA implementados en el Distrito deberán garantizar la ausencia de sesgos discriminatorios basados en categorías protegidas por la normativa colombiana, incluyendo raza, género, orientación sexual, discapacidad y condición socioeconómica. Este compromiso se materializa mediante la aplicación de pruebas obligatorias de equidad en todas las fases del ciclo de vida de los sistemas.

La implementación incluye la definición de umbrales máximos de disparidad específicos para cada tipo de servicio, asegurando que los sistemas no repliquen o amplifiquen patrones de discriminación existentes. Este enfoque preventivo garantiza que la IA promueva la igualdad de trato en la prestación de servicios distritales.

Seguridad y robustez. Los sistemas de IA implementados en el Distrito deberán garantizar niveles adecuados de seguridad y resistencia técnica ante fallos operativos y amenazas cibernéticas. Para ello, se implementarán controles de seguridad basados en el marco del NIST Cybersecurity Framework, adaptados a las particularidades de los sistemas de inteligencia artificial.

Como parte integral del proceso de validación, se realizarán pruebas de robustez que incluirán escenarios de estrés operativo, evaluación con datos atípicos y simulaciones de intentos de manipulación. Estas verificaciones asegurarán que los sistemas mantengan su funcionalidad y confiabilidad incluso en condiciones adversas, protegiendo la integridad de los servicios distritales.

Privacidad por diseño y por defecto. La protección de datos personales se integrará desde la fase inicial de diseño de todos los sistemas de IA, aplicando configuraciones predeterminadas que maximicen la privacidad de los ciudadanos. Este enfoque garantiza el cumplimiento de la Ley 1581 (2012) de 2012 y la Circular Externa 002 de 2024 de la SIC mediante la implementación sistemática de medidas técnicas y organizativas.

La implementación incluirá técnicas de minimización de datos, seudonimización y anonimización, junto con controles de acceso basados en el principio de privilegio mínimo. Estas medidas asegurarán que el tratamiento de información personal se realice dentro de los marcos legales establecidos, priorizando la protección de los derechos de los titulares en todas las operaciones con sistemas de IA.

La Política de Uso de IA complementa estos principios con directrices operativas específicas para diferentes tipos de casos de uso, estableciendo requisitos diferenciados según el nivel de riesgo y el dominio de aplicación (atención ciudadana, trámites administrativos, analítica para políticas públicas, gestión de recursos, salud pública, movilidad).

4.1.1.1. Política de Gobierno de Datos

La Política de Gobierno de Datos establece los estándares y procedimientos para garantizar la calidad, integridad y gestión adecuada de los datos utilizados en sistemas de inteligencia artificial del Distrito. Reconociendo que los datos constituyen el fundamento crítico para el éxito y la confiabilidad de estas tecnologías, esta política regula todo el ciclo de vida de la información, desde su obtención hasta su disposición final.

La implementación de esta política asegura que los datos cumplan con requisitos de representatividad, exactitud y actualidad, mitigando riesgos asociados a sesgos y errores. Este marco permite a las entidades distritales contar con información confiable para la toma de decisiones automatizadas, manteniendo el cumplimiento de la normativa vigente en protección de datos personales.

La Figura 2 sintetiza los componentes fundamentales de la Política de Gobierno de Datos propuesta, incluyendo estándares de calidad y representatividad, principios de minimización y propósito específico, mecanismos de gestión de sesgos y discriminación, la obligatoriedad de evaluaciones de impacto en privacidad (ARA/DPIA), garantías para el ejercicio de los derechos de los titulares y criterios de localización y soberanía de datos.

Figura 2 *Política de Gobierno de Datos: Componentes Principales*

Fuente: Elaboración propia (2025).

Los componentes principales incluyen:

Calidad y representatividad de datos. Se establecen estándares mínimos de calidad que garantizan completitud, precisión, consistencia, actualidad y validez de los datos. Para casos sensibles, se requiere análisis estadístico de representatividad demográfica y documentación de limitaciones mediante Data Sheets, asegurando que los datos reflejen adecuadamente la diversidad poblacional de Bogotá.

Minimización y propósito específico. En cumplimiento de la Ley 1581 (2012), se recolectarán únicamente los datos estrictamente necesarios para el propósito declarado, evitando acumulación innecesaria. Se prohíbe el uso de datos para fines incompatibles sin la debida autorización legal.

Gestión de sesgos y discriminación. Se implementa evaluación sistemática de sesgos en datos fuente, con documentación obligatoria de sesgos identificados y aplicación de técnicas de mitigación cuando

sea viable. Se reconoce que no todos los sesgos son eliminables, requiriéndose transparencia sobre limitaciones residuales.

Evaluaciones de impacto en privacidad. Siguiendo las directrices de la SIC Circular Externa 002/2024, se establece la obligatoriedad de realizar Análisis de Riesgos y Gestión (ARA) o Data Protection Impact Assessments (DPIA) para todo sistema de IA que procese datos personales, con especial profundidad para datos sensibles o tratamientos de alto riesgo utilizando la plantilla DPIA del toolkit (sección 4.2.3).

Derechos de los titulares. Se establecen mecanismos para garantizar el ejercicio efectivo de los derechos de habeas data (conocer, actualizar, rectificar, suprimir, revocar) en el contexto de sistemas de IA. Para sistemas con capacidades de aprendizaje continuo, se documentan las implicaciones técnicas del derecho al olvido y se implementan procesos de "desaprendizaje" cuando sea técnicamente viable.

Localización y soberanía de datos. Para datos sensibles o críticos, se establecen requisitos de localización en infraestructura nacional o con garantías contractuales estrictas de jurisdicción y protección legal. En casos de transferencia internacional de datos, se exige el cumplimiento de los requisitos del Capítulo IV del Decreto 1377 de 2013 (Presidencia de la República, 2013) y cláusulas contractuales estándar que garanticen un nivel de protección adecuado.

4.1.1.2. Política de Gestión de Riesgos de IA

Esta política establece un marco integral para la identificación, evaluación y gestión de riesgos asociados a los sistemas de IA en el Distrito, integrando estándares internacionales (NIST AI RMF 2023, ISO/IEC 23894) con el enfoque de clasificación por niveles del AI Act, adaptado al contexto regulatorio colombiano.

Sistema de clasificación de riesgos: Adoptando el enfoque basado en riesgo del AI Act, se establece una taxonomía de cuatro niveles:

Riesgo Inaceptable (Prohibido). Categoría que incluye sistemas que vulneran derechos fundamentales o principios constitucionales. En el ámbito distrital, comprende: sistemas de puntuación social ciudadana, manipulación subliminal, causante de daño, explotación de vulnerabilidades de grupos específicos, biometría remota en tiempo real para vigilancia masiva sin orden judicial.

Alto Riesgo. Sistemas con impacto significativo en derechos fundamentales, seguridad o acceso a servicios esenciales, que requieren obligaciones reforzadas de evaluación, documentación y auditoría. Incluye: evaluación y priorización de beneficiarios de programas sociales, asistencia en decisiones judiciales o administrativas, sistemas biométricos de identificación, gestión de infraestructura crítica (movilidad, servicios públicos), evaluación de elegibilidad para servicios de salud o educación.

Riesgo Limitado. Sistemas con requisitos de transparencia básicos, como chatbots de atención ciudadana, recomendación de información pública y asistentes virtuales para trámites no sensibles. La obligación principal es la divulgación clara de la interacción con IA.

Riesgo Mínimo. Sistemas sin impacto significativo en derechos o servicios esenciales, como filtros de spam, optimización de rutas internas y herramientas de productividad. Aplican principios generales sin evaluaciones formales obligatorias.

Este enfoque proporcional permite una implementación eficiente de controles, focalizando recursos en los sistemas de mayor impacto y riesgo.

Metodología de evaluación de riesgos: Basada en el marco del NIST AI RMF, esta política establece un proceso estructurado en cuatro fases para la gestión integral de riesgos en sistemas de IA:

Identificación de riesgos. Identificación y análisis sistemático que considera el contexto de uso, actores afectados, datos utilizados y tipo de decisiones, identificando riesgos potenciales en categorías clave como sesgos, privacidad, seguridad, explicabilidad y robustez técnica.

Medición de riesgos. Evaluación cuantitativa y cualitativa de probabilidad e impacto mediante matrices estandarizadas, considerando dimensiones críticas como afectación a derechos fundamentales, equidad, continuidad del servicio, seguridad y reputación institucional.

Gestión de riesgos. Selección y aplicación de medidas proporcionales al nivel de riesgo identificado, que incluyen prevención, detección, mitigación, transferencia o aceptación de riesgos, documentadas en planes de mitigación específicos.

Gobernanza de riesgos. Asignación clara de responsabilidades, procesos de escalamiento, y mecanismos de monitoreo continuo con revisiones periódicas y activación de protocolos de respuesta ante materialización de riesgos.

4.1.1.3. Política de Compras y Proveedores de IA

Ante la dependencia identificada del Distrito de proveedores externos para soluciones de inteligencia artificial (identificada en el diagnóstico de la Fase 1), esta política establece requisitos de debida diligencia y cláusulas contractuales mínimas para mitigar riesgos asociados a la cadena de suministro tecnológica.

El objetivo es mitigar los riesgos asociados a la cadena de suministro tecnológica mediante la estandarización de procesos de selección y contratación, asegurando que los proveedores cumplan con los estándares éticos, técnicos y normativos del Distrito.

Los requisitos de conformidad regulatoria establecen que los proveedores deben demostrar cumplimiento con: el marco regulatorio colombiano aplicable como la Ley 1581 (2012), normativa SIC, CONPES 4144 (2024); para sistemas de alto riesgo, la conformidad con obligaciones equivalentes a las del AI Act o compromiso de certificación dentro de plazos razonables; las políticas públicas de IA responsable y transparencia, preferiblemente con informes anuales verificables; y un Sistema de Gestión de IA (AIMS) según ISO/IEC 42001 implementado o en vías de certificación, con evidencia documentada.

En cuanto a los requisitos de documentación técnica, los proveedores deberán entregar documentación técnica completa que garantice la transparencia y auditabilidad de los sistemas de IA. Esta incluye: Model Cards describiendo arquitectura, propósito, datos de entrenamiento, métricas de desempeño, limitaciones conocidas y casos de uso apropiados/inapropiados; Data Sheets para todos los conjuntos de datos utilizados, documentando metodología de recolección, composición demográfica, procesos de limpieza y limitaciones; documentación técnica detallada para auditoría (arquitectura, hiperparámetros, proceso de entrenamiento, resultados de pruebas de robustez y equidad); y bitácoras de decisiones de diseño y gestión de sesgos. Esta documentación permitirá al Distrito verificar la calidad técnica, comprender las limitaciones operativas y mantener la trazabilidad necesaria para la auditoría continua de los sistemas implementados.

Los requisitos de privacidad y seguridad incluyen: claridad sobre la titularidad de datos (distinguiendo datos de entrenamiento, operación y metadatos generados); especificación de localización física y jurisdiccional de datos; políticas de retención, borrado y portabilidad de datos con compromisos contractuales exigibles; Data Processing Agreements (DPA) para tratamiento de datos personales, estableciendo roles de responsable y encargado según Ley 1581 (2012); certificaciones de seguridad relevantes (ISO 27001, SOC 2) y resultados de pruebas de robustez y pentesting; y compromiso de notificación de incidentes de seguridad dentro de plazos definidos. Estos requisitos aseguran que los proveedores manejen los datos del Distrito con los más altos estándares de seguridad y en estricto cumplimiento de la normativa colombiana de protección de datos personales.

Los requisitos de auditoría y mejora continua establecen que los contratos con proveedores de sistemas de IA incorporarán cláusulas específicas que garanticen la supervisión continua y la evolución controlada de las soluciones implementadas. Estas incorporan: cláusulas de derecho a auditoría, permitiendo evaluaciones independientes con acceso a documentación, código (cuando sea viable), y evidencias de pruebas; implementación de sistemas de telemetría para detectar desviaciones en el comportamiento de los modelos (drift), degradación de desempeño y sesgos emergentes; Acuerdos de Niveles de Servicio (Service Level Agreements) con compromisos de disponibilidad, tiempo de

respuesta, y calidad del servicio; y un roadmap de producto, gestión de versiones y control de cambios con procesos de notificación y evaluación de impacto de actualizaciones. Estos requisitos permiten al Distrito mantener supervisión efectiva sobre los sistemas de IA a lo largo de su ciclo de vida, asegurando que evolucionen de manera controlada y mantengan los estándares de calidad y ética establecidos.

El Checklist de Evaluación de Proveedores del toolkit (sección 4.2.5) operacionaliza estos requisitos en un instrumento práctico para procesos de selección y contratación.

4.1.2. Capa 2: Modelo de Gobierno

Esta capa establece la estructura organizacional y los mecanismos de decisión para implementar la gobernanza de IA en el Distrito Capital, reconociendo las diferentes capacidades institucionales. El modelo ofrece dos modalidades prácticas: un Comité Distrital centralizado para entidades con menor madurez técnica, y Comités por entidad para organizaciones con mayor capacidad y volumen de casos de uso.

El diseño incorpora mecanismos claros de toma de decisiones y se integra con los sistemas de gestión existentes, incluyendo el Modelo Integrado de Planeación y Gestión (MIPG) -marco de referencia para la gestión pública en Colombia- y el control interno, asegurando que la gobernanza de IA se implemente de manera articulada con la operación institucional. La composición de los comités refleja roles realistas y representativos, facilitando la implementación efectiva del Framework en toda la administración distrital.

4.1.2.1. Comité de IA Distrital o por Entidad

El Comité de IA se establece como la máxima instancia de gobernanza para la inteligencia artificial en el Distrito Capital, responsable de la dirección estratégica, supervisión y coordinación de todas las iniciativas en esta materia. Como órgano colegiado de carácter multidisciplinario, integra las perspectivas técnicas, jurídicas, éticas y operativas necesarias para una toma de decisiones balanceada y fundamentada. Su diseño operativo incorpora mecanismos claros de toma de decisiones y se articula con los sistemas de gestión existentes, incluyendo el MIPG y el control interno, asegurando que la gobernanza de IA se implemente de manera coherente con la estructura institucional del Distrito. La composición del comité refleja roles realistas y representativos, facilitando la implementación efectiva del Framework en toda la administración distrital.

La composición y roles del Comité se estructuran de la siguiente manera: la Alta Dirección (Presidente del Comité) corresponde a un Secretario, Director o nivel directivo equivalente que aporta patrocinio institucional, alineación estratégica con objetivos de la entidad y capacidad de decisión sobre recursos

y prioridades. El Oficial de IA Responsable (Secretaría Técnica) es un rol dedicado responsable de la coordinación operativa del comité, seguimiento de casos de uso, gestión de riesgos corporativos de IA y enlace con la Alta Dirección, pudiendo ser desempeñado por el responsable de transformación digital, innovación o calidad según la estructura organizacional. El Representante de Tecnologías de Información (TIC) y Seguridad aporta perspectiva técnica sobre arquitectura, desarrollo, operación, ciberseguridad y viabilidad de implementación, siendo responsable de evaluar requisitos técnicos y garantizar la robustez de los sistemas. El Data Protection Officer (DPO) o Responsable de Habeas Data asegura el cumplimiento de la Ley 1581 (2012) y normativa SIC, evalúa ARA/DPIA, supervisa derechos de los titulares y gestiona riesgos de privacidad, contando con voto vinculante en decisiones que afecten datos personales. El Representante Jurídico evalúa bases legales para tratamiento de datos, conformidad regulatoria, viabilidad de contratos con proveedores y gestión de riesgos legales, siendo responsable de la revisión de cláusulas contractuales y términos de servicio. El Representante de Planeación asegura la alineación con el plan estratégico institucional, evalúa viabilidad presupuestal, coordina con otros proyectos de transformación y vela por la sostenibilidad de iniciativas. El Representante de Control Interno aporta perspectiva de gestión de riesgos institucionales, evalúa controles y mecanismos de auditoría, y vela por la coherencia con sistemas de gestión de calidad y control interno existentes. El Representante de Atención al Ciudadano o Área de Negocio aporta comprensión profunda del contexto de uso, necesidades de usuarios finales, viabilidad operativa y calidad del servicio, asegurando que las soluciones respondan a problemáticas reales y sean apropiadas para el perfil de los ciudadanos beneficiarios.

Las funciones del Comité incluyen: aprobar o rechazar la implementación de casos de uso de IA basándose en la evaluación de riesgos, conformidad normativa y alineación estratégica; supervisar el portafolio de iniciativas de IA, priorizando según impacto, riesgo y recursos disponibles; establecer y actualizar políticas institucionales de IA alineadas con el framework distrital; revisar y aprobar ARA/DPIA para casos de alto riesgo; monitorear KPIs agregados y gestionar incidentes significativos; aprobar contratos con proveedores de IA y renovaciones basándose en evaluación de desempeño; coordinar capacitación y gestión del cambio organizacional; supervisar la implementación del programa obligatorio de capacitación en IA Generativa**;** y mantener enlace con instancias distritales superiores (para Comités por entidad) o con otras entidades (para Comité Distrital centralizado).

Respecto a las modalidades de implementación, se contemplan dos opciones. El Comité de IA Distrital (centralizado) es recomendado para la fase inicial de implementación o para entidades de menor tamaño y madurez, donde un único Comité a nivel del Distrito brinda servicios de gobernanza a múltiples entidades, asegurando consistencia de criterios, compartiendo costos de expertise especializada y facilitando el aprendizaje inter-institucional, requiriendo la designación de un equipo

técnico de soporte con capacidad para atender múltiples entidades. Los Comités por Entidad son recomendados para entidades de gran tamaño, alta complejidad o con volumen significativo de casos de uso (más de 5 sistemas en operación o planificación), permitiendo mayor agilidad en la toma de decisiones, especialización en el contexto específico de la entidad y apropiación institucional más profunda, requiriendo coordinación con el nivel distrital para consistencia normativa y compartición de lecciones aprendidas.

4.1.2.2. Roles a Nivel de Caso de Uso

Complementando la estructura de gobernanza corporativa, se definen roles operativos específicos para cada iniciativa o sistema de IA.

El Sponsor de Negocio (Product Owner) es el líder del área usuaria que impulsa la iniciativa de IA, definiendo su propósito estratégico y requisitos funcionales. Como responsable último de los resultados, asegura la alineación con los objetivos institucionales, gestiona el cambio organizacional y garantiza la apropiación del sistema en su área, rindiendo cuentas sobre su impacto operativo.

El Responsable Técnico (Technical Lead) lidera el desarrollo o implementación técnica, asegura la calidad del sistema, supervisa pruebas, coordina con proveedores externos y es responsable del monitoreo técnico continuo. Este rol debe tener conocimientos sólidos en IA/ML y comprensión de los principios de IA responsable.

El Propietario de Datos (Data Steward) gestiona integralmente el ciclo de vida de los datos utilizados por el sistema, garantizando su calidad, disponibilidad y gobierno. En coordinación con el Oficial de Datos Personales y el área técnica, vela por el cumplimiento normativo de privacidad y la correcta gestión de los datos desde su obtención hasta su disposición final.

Los Usuarios Finales y Representantes de Ciudadanía participan activamente en el diseño, validación y mejora continua del sistema, aportando perspectivas basadas en la experiencia real de uso. En casos de alto riesgo, su participación se formaliza mediante grupos focales que aseguran que las voces de las comunidades afectadas sean consideradas.

La matriz RACI (Responsible, Accountable, Consulted, Informed) del toolkit especifica las responsabilidades detalladas para cada rol en las diferentes fases del ciclo de vida.

4.1.2.3. Matriz de Responsabilidades (RACI)

Para garantizar claridad en la ejecución del framework, se establece la siguiente matriz RACI que define roles por fase del ciclo de vida:

Tabla 3. Capa 2: Matriz de Responsabilidad (RACI)

Fase del Ciclo de Vida	Comité de IA	DPO	Responsable Técnico	Sponsor de Negocio	Área Jurídica
Intake y Canvas	C	C	R	A	C
Clasificación de Riesgo	A	C	R	I	C
ARA/DPIA	A	A (voto vinculante)	C	C	C
Gobierno de Datos	C	A	R	C	I
Desarrollo/Adquisición	C	C	A	C	A (contratos)
Pruebas y Validación	C	C	A	C	I
Despliegue	A (go-live)	C	R	C	I
Monitoreo y Auditoría	A	C	R	I	C
Retiro	A	A (datos)	R	C	C

Fuente: Elaboración Propia

El DPO ejerce veto vinculante en fases que involucren tratamiento de datos personales (ARA/DPIA, Gobierno de Datos, Retiro). El Comité de IA actúa como gate decisorio en todas las fases críticas, mientras que la Alta Dirección retiene autoridad final para sistemas de alto riesgo o inversiones significativas.

4.1.2.4. Instrumentos Operativos de Gobernanza

Para facilitar la coordinación, trazabilidad y eficiencia en la gestión del portafolio de IA, se establecen dos instrumentos operativos obligatorios:

Registro Central de Casos de Uso de IA

Sistema centralizado que documenta todos los casos de uso en operación, desarrollo o evaluación. Estructura mínima de datos: ID único, nombre del sistema, entidad responsable, clasificación de riesgo, estado del ciclo de vida, fecha de inicio, responsables (técnico, sponsor, DPO), descripción breve, datos

personales tratados (Sí/No/Sensibles), fecha de última actualización. Actualización obligatoria: mensual para sistemas en operación, semanal para sistemas en desarrollo. Acceso: consulta pública (datos no sensibles) para transparencia ciudadana; acceso completo para Comité de IA, control interno y auditoría. Responsable: Oficial de IA Responsable (Secretaría Técnica del Comité).

Catálogo Distrital de Proveedores Pre-evaluados

Listado de proveedores que han superado evaluación previa mediante el Checklist de Evaluación (sección 4.2.6), clasificados por nivel de riesgo y tipo de solución. Criterios de inclusión: aprobación ≥75% en checklist estándar, cumplimiento obligatorio de Ley 1581 (2012) y Ley 2279 (2022), certificaciones de seguridad vigentes, referencias verificables en sector público. Beneficios: reducción de 40-60% en tiempo de evaluación para licitaciones, estandarización de cláusulas contractuales, mayor poder de negociación institucional. Actualización: anual con revisión semestral de certificaciones. Responsable: Comité de IA con soporte del área de Contratación.

Estos instrumentos se integran al dashboard de gobernanza (Anexo C) mediante indicadores de cobertura de registro (objetivo: 100%) y uso del catálogo en procesos de contratación (objetivo: ≥70% para sistemas de alto riesgo)

4.1.2.5. Alineación con CONPES 4144 y Vigilancia de Cumplimiento

El modelo de gobierno establece mecanismos específicos de trazabilidad y reporte que garantizan la alineación estratégica con los cinco ejes del marco nacional (CONPES, 2024), integrando la gobernanza de IA con los sistemas de gestión existentes en el Distrito.

El Eje de Ética y Gobernanza establece que el Comité de IA supervisa la implementación de la Carta de IA Responsable y reporta trimestralmente el estado del portafolio, incidentes éticos y lecciones aprendidas, integrando estos indicadores al Sistema de Seguimiento a Metas del MIPG.

En el Eje de Datos e Infraestructura, el DPO y el área TIC reportan trimestralmente sobre la madurez de la gobernanza de datos, calidad de conjuntos de datos utilizados y estado de la infraestructura tecnológica, información que se incorpora a los indicadores de gestión del MIPG.

El Eje de Gestión de Riesgos mantiene un registro corporativo de riesgos de IA actualizado trimestralmente, el cual se articula con la matriz de riesgos del sistema de control interno, reportando riesgos materializados, efectividad de controles y necesidades de mejora.

Respecto al Eje de Talento, el área de Gestión Humana reporta semestralmente los avances del programa de capacitación en IA responsable, cobertura de formación y desarrollo de capacidades internas, integrando estas métricas al plan de desarrollo del talento humano.

En el Eje de Uso y Adopción, se cuantifica y reporta trimestralmente el avance en adopción mediante indicadores de casos de uso por fase del ciclo de vida, servicios transformados, beneficiarios impactados y resultados en eficiencia y calidad del servicio.

Para la vigilancia de cumplimiento regulatorio, el Comité establece un calendario de revisiones que incluye: cumplimiento de Ley 1581 (2012) y normativa SIC, con revisión mensual de ARA/DPIA vigentes, atención de derechos de habeas data, y respuesta a requerimientos de la autoridad de protección de datos; la conformidad con políticas de gobierno digital y transparencia, mediante la publicación de inventario de sistemas de IA (cuando aplique según normativa de datos abiertos), divulgaciones ciudadanas requeridas, y rendición de cuentas; y el seguimiento al plan de acción derivado del CONPES 4144 (2024) a nivel institucional con reporte anual de avances al nivel distrital correspondiente.

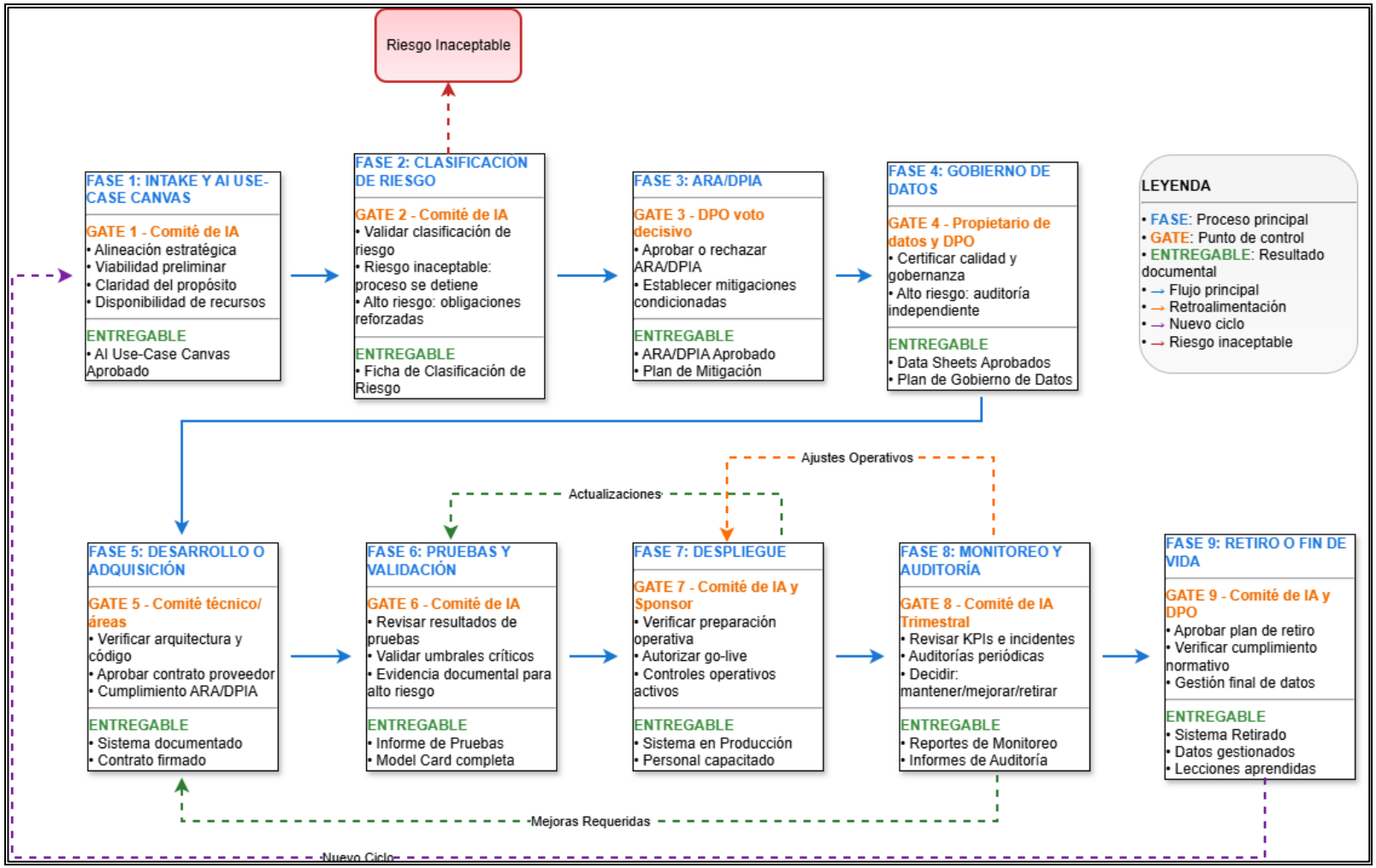
Esta estructura asegura que la implementación del Framework de IA mantenga coherencia total con la política nacional, mientras se integra eficientemente con los sistemas de gestión y control ya establecidos en la administración distrital.

4.1.3. Capa 3: Fases del Framework de Gobernanza de IA

El modelo propuesto se organiza en nueve etapas sucesivas que pretenden garantizar una adopción responsable y estructurada de sistemas de inteligencia artificial. Cada fase incorpora controles progresivos que reducen los riesgos desde la concepción del proyecto hasta su retiro definitivo.

La Figura 3 presenta el Proceso de Ciclo de Vida de Gobernanza de IA, estructurado como un diagrama de flujo que integra las fases secuenciales, los puntos de control (gates) y los entregables obligatorios en cada etapa. Este modelo permite visualizar de forma clara cómo la gobernanza se articula desde la concepción del caso de uso, pasando por la evaluación de riesgos y la validación técnica, hasta la puesta en producción, seguimiento y auditoría continua. Asimismo, evidencia la estructura de toma de decisiones y los hitos de revisión que garantizan trazabilidad, transparencia y control institucional en todo el ciclo de vida de los sistemas de IA, alineándose con la arquitectura de cinco capas del Framework propuesto.

Figura 3 ***Proceso de Ciclo de Vida de Gobernanza de IA***



Fuente: Elaboración Propia

4.1.3.1. Fase 1 – Intake y AI Use-Case Canvas

En la fase inicial se formaliza la idea del caso de uso mediante el *AI Use-Case Canvas* (disponible en el toolkit, sección 4.2.1). El sponsor de negocio completa la plantilla describiendo el problema, los objetivos, los actores involucrados, los datos requeridos, los resultados esperados y las métricas de éxito. Simultáneamente, el responsable técnico, el oficial de protección de datos (DPO) y el área de planeación realizan una primera valoración de viabilidad técnica, legal y presupuestal, al tiempo que identifican riesgos y partes interesadas.

El Comité de IA actúa como *gate* de revisión (G1). Sus criterios –alineación estratégica, viabilidad preliminar, claridad del propósito y disponibilidad de recursos– determinan si el caso avanza a la clasificación de riesgo. El entregable de esta fase es el *AI Use-Case Canvas* aprobado.

4.1.3.2. Fase 2 – Clasificación de Riesgo

Esta etapa asigna un nivel de riesgo (inaceptable, alto, limitado o mínimo) siguiendo la taxonomía del AI Act. La matriz de clasificación evalúa propósito del sistema, tipo de decisiones, datos procesados, población afectada, reversibilidad y consecuencias de errores. Cuando el caso se ubica en la categoría prohibida (riesgo inaceptable), el proceso se detiene; para los de alto riesgo se activan obligaciones reforzadas (ARA/DPIA obligatorio, documentación técnica ampliada, supervisión humana y auditoría).

El Comité de IA revisa y valida la clasificación (G2). El resultado se plasma en la *Ficha de Clasificación de Riesgo*, que constituye el documento de referencia para las fases posteriores.

4.1.3.3. Fase 3 – ARA/DPIA (alto riesgo o datos sensibles)

El objetivo es identificar y mitigar impactos sobre derechos fundamentales y la privacidad antes de la puesta en marcha. Se completa la *Plantilla ARA/DPIA* (toolkit, sección 4.2.3), que incluye: descripción del sistema, mapeo de flujos de datos, bases legales de tratamiento, evaluación de impactos en privacidad, no discriminación, debido proceso y acceso equitativo; análisis de riesgos mediante la *Matriz de Riesgos* (sección 4.2.2); y propuestas de medidas técnicas y organizativas (minimización, seudonimización, cifrado, auditoría, supervisión humana).

El DPO, como voto decisivo dentro del Comité de IA, aprueba o rechaza el documento (G3). Cuando la aprobación está condicionada, se establecen mitigaciones específicas que deben incorporarse antes de avanzar. El entregable es el ARA/DPIA aprobado con su plan de mitigación.

4.1.3.4. Fase 4 – Gobierno de Datos

En esta fase se asegura que los datos que alimentarán el modelo sean de alta calidad, representativos y gestionados éticamente. Las actividades comprenden: inventario y documentación de fuentes (internas, externas, terceros); evaluación de calidad dimensional (completitud, precisión, consistencia, actualidad) mediante la lista de verificación del toolkit; análisis de representatividad demográfica y detección de sesgos; aplicación de técnicas de mitigación (remuestreo, reponderación, recolección complementaria); y elaboración de *Data Sheets* siguiendo la plantilla de Gebru et al. (2018).

El propietario de datos y el DPO certifican la calidad y la gobernanza adecuada (G4). En casos de alto riesgo, se puede requerir auditoría independiente o validación estadística formal. El resultado son los *Data Sheets* aprobados y el plan de gobernanza de datos.

4.1.3.5. Fase 5 – Desarrollo o Adquisición

Esta etapa traduce los requisitos de gobernanza, seguridad y calidad en un producto técnico.

Desarrollo interno. Se convierten los requisitos funcionales y no funcionales (incluidos los de IA responsable) en especificaciones técnicas; se elige arquitectura, algoritmos y tecnologías que favorezcan la interpretabilidad sin sacrificar el desempeño crítico; se incorporan controles de privacidad por diseño (minimización, seudonimización, control de accesos) y de seguridad (validación de entradas, defensa contra ataques adversarios, cifrado). Además, se implementa versionado de código, datos de entrenamiento y modelos para garantizar trazabilidad, y se mantiene una documentación continua de decisiones de diseño, hiperparámetros y procesos de entrenamiento.

Adquisición externa. Se redactan términos de referencia que integran el *Checklist de Proveedores* (toolkit, sección 4.2.5); se evalúan propuestas ponderando criterios técnicos, de conformidad regulatoria y experiencia del proveedor; y se negocian contratos que incluyan cláusulas de privacidad (DPA), seguridad, auditoría, SLA y gestión de cambios. La documentación entregada por el proveedor tales como Model Cards, Data Sheets (toolkit, sección 4.2.4) y certificaciones sen valida antes de la firma.

El *gate* de revisión (G5) difiere según la vía elegida: para desarrollo interno, el comité técnico verifica arquitectura y código; para adquisición, el área jurídica, el DPO y el Comité de IA aprueban el contrato y confirman que el sistema satisface el ARA/DPIA aprobado (toolkit, sección 4.2.3). El entregable es el sistema (con documentación técnica completa) o el contrato firmado.

4.1.3.6. Fase 6 – Pruebas y Validación

Antes del despliegue, el sistema debe demostrar cumplimiento con requisitos funcionales, técnicos, éticos y de usabilidad. Las pruebas se agrupan en cinco categorías.

Las pruebas técnicas evalúan métricas objetivas (precisión, recall, F1, AUC ROC), utilizan conjuntos de validación independientes y examinan la robustez frente a datos fuera de distribución, ruido y pequeñas perturbaciones. Además, se realizan pruebas de seguridad (penetration testing) y de rendimiento (tiempo de respuesta, uso de recursos, escalabilidad).

Las pruebas de equidad calculan métricas de fairness (disparate impact, equal opportunity, predictive parity) y desglosan los resultados por grupos demográficos relevantes. Cuando aparecen disparidades inaceptables, se aplican técnicas de mitigación (post processing, optimización de umbrales) y se documentan los trade offs entre equidad y desempeño.

Las pruebas de explicabilidad verifican la presencia de mecanismos explicativos (LIME, SHAP, mapas de atención, contra factuales) y validan su comprensibilidad mediante pruebas con usuarios finales.

Las pruebas de usabilidad y accesibilidad consisten en la realización de sesiones con usuarios representativos (funcionarios, ciudadanos) siguiendo los principios de diseño centrado en el usuario y aseguran el cumplimiento de WCAG 2.1 nivel AA para personas con discapacidad.

Las pruebas de integración comprueban la interoperabilidad con sistemas legados, la correcta exposición de APIs y la capacidad de recuperación ante fallos.

El responsable técnico presenta los resultados al Comité de IA (G6). Para sistemas de alto riesgo, la aprobación depende de evidencia documental que demuestre el cumplimiento de umbrales predefinidos en desempeño, equidad, robustez y seguridad. El entregable es el Informe de pruebas y validación acompañado de una Model Card completa (toolkit, sección 4.2.6).

4.1.3.7. Fase 7 – Despliegue

El objetivo es poner el sistema en producción de forma controlada, garantizando que el personal esté preparado y que los controles operativos estén activos. Las actividades incluyen: capacitación de usuarios finales (operadores y, cuando corresponda, ciudadanos) sobre funcionamiento, limitaciones y procedimientos de supervisión humana; configuración de logs de auditoría, telemetría, alertas y dashboards de monitoreo; despliegue gradual (piloto limitado seguido de escalamiento progresivo); establecimiento de canales de reporte de incidentes y quejas; y comunicación transparente a la ciudadanía cuando el sistema interactúe con el público.

El Comité de IA verifica la preparación operativa (G7) y el sponsor de negocio autoriza el *go-live*. El entregable es el sistema en producción con los controles operativos activos y el personal debidamente capacitado.

4.1.3.8. Fase 8 – Monitoreo y Auditoría

Una vez en funcionamiento, el sistema requiere supervisión continua para detectar desviaciones y aplicar correcciones. Las actividades se estructuran en cuatro áreas.

El monitoreo técnico continuo comprende el seguimiento de métricas de desempeño, la detección de data drift y concept drift, el cálculo periódico de métricas de equidad y la vigilancia de intentos de ataque.

La gestión de incidentes implica el registro, clasificación y respuesta a incidentes (técnicos, de privacidad, de equidad, de seguridad o quejas ciudadanas) siguiendo protocolos de contención, investigación, remediación y comunicación.

La auditoría periódica consiste en la revisión trimestral (o con mayor frecuencia según el nivel de riesgo) del cumplimiento de los controles del ARA/DPIA, la auditoría de logs para validar la supervisión humana y el análisis de quejas. Los sistemas de alto riesgo requieren una auditoría anual independiente que evalúe la conformidad técnica, ética y regulatoria.

La gestión de cambios abarca la evaluación de impacto de actualizaciones (nuevas versiones del modelo, cambios en datos o código), la re-ejecución de pruebas críticas y la actualización de la documentación (Model Cards, Data Sheets).

El Comité de IA revisa trimestralmente el dashboard de KPIs, incidentes y hallazgos de auditoría (G8) y decide si se mantiene la operación, se implementan mejoras o se procede al retiro. Los entregables son los reportes de monitoreo, los registros de incidentes y los informes de auditoría.

4.1.3.9. Fase 9 – Retiro o Fin de Vida

Cuando el sistema deja de ser necesario, viable o conforme, se lleva a cabo un retiro ordenado que preserve la continuidad del servicio y garantice una gestión adecuada de los datos. Las actividades incluyen: decisión de retiro basada en obsolescencia técnica, cambios regulatorios, riesgos inaceptables, fin del propósito o análisis costo-beneficio desfavorable; planificación de la transición (retorno a procesos manuales o sustitución por otro sistema); gestión de datos (exportación para fines legales, anonimización o borrado seguro conforme a la Ley 1581 (2012)); documentación de lecciones aprendidas y comunicación a los ciudadanos afectados.

El Comité de IA aprueba el plan de retiro y el DPO verifica el cumplimiento de la normativa de protección de datos (G9). El entregable final es el sistema retirado, los datos gestionados conforme a la normativa y un informe de lecciones aprendidas archivado.

4.1.4. Capa 4: Controles Clave

Esta capa define los controles técnicos y organizacionales necesarios para implementar de manera transversal los principios éticos y mitigar riesgos a lo largo del ciclo de vida de los sistemas de inteligencia artificial. Los controles se estructuran en seis dimensiones fundamentales.

4.1.4.1. Ética y Derechos Fundamentales

Supervisión humana significativa. En decisiones que afecten derechos fundamentales o servicios esenciales, se requieren mecanismos que garanticen una supervisión humana competente. Esta supervisión debe permitir comprender el funcionamiento y las limitaciones del sistema, monitorear su operación en tiempo real o revisar decisiones *ex-post* según la criticidad, intervenir o anular decisiones algorítmicas cuando sea necesario, y activar protocolos de escalamiento ante situaciones fuera del diseño original. La supervisión debe ser genuina—ya sea *human-in-the-loop* u *human-on-the-loop*—y no meramente ceremonial, lo que exige interfaces adecuadas y capacitación específica del personal supervisor.

No discriminación y equidad. Los controles implementados incluyen análisis de sesgos en los datos de entrenamiento y la aplicación de técnicas de mitigación, pruebas de equidad durante el desarrollo y monitoreo continuo de métricas de *fairness*. Se establecen umbrales máximos de disparidad según el tipo de servicio; por ejemplo, para servicios sociales críticos se podría requerir un *disparate impact ratio* no inferior a 0.8. Además, se incorporan mecanismos de explicación que permitan detectar discriminación en casos individuales y procesos de apelación accesibles cuando las decisiones algorítmicas afecten negativamente a los ciudadanos.

Acceso equitativo. El diseño debe ser inclusivo, considerando la accesibilidad para personas con discapacidad visual, auditiva, motriz o cognitiva. Es crucial disponer de canales alternativos de atención humana para quienes no puedan utilizar el sistema de IA, así como adaptar las interfaces a diversos niveles de alfabetización digital. Finalmente, se debe evitar condicionar el acceso a servicios esenciales exclusivamente a sistemas digitales, mitigando así la brecha digital.

4.1.4.2. Privacidad y Protección de Datos

Bases legales claras. Todo tratamiento de datos personales debe fundamentarse en una base legal válida conforme a la Ley 1581 (2012), como el consentimiento informado, previo y expreso; la ejecución de un contrato; el cumplimiento de una obligación legal; la protección de intereses vitales;

o el ejercicio de funciones públicas. El Análisis de Riesgos y Aspectos (ARA) o la Evaluación de Impacto en la Protección de Datos (DPIA) deben documentar específicamente la base legal aplicable.

Minimización de datos. Se aplica el principio de que solo deben recolectarse y procesarse los datos estrictamente necesarios para el propósito declarado. Esto implica una evaluación crítica de cada campo de datos, preferencia por datos agregados o anonimizados cuando sea posible, evitar la recolección especulativa y realizar revisiones periódicas para eliminar datos innecesarios.

Anonimización y seudonimización. Se emplean técnicas de protección según la proporcionalidad: la anonimización irreversible para datos estadísticos y la seudonimización cuando se requiera trazabilidad sin identificación rutinaria. Es esencial evaluar el riesgo de re-identificación y reforzar los controles de acceso para las llaves de des-seudonimización.

DPIA obligatorio. Para tratamientos de alto riesgo según la Circular 002/2024 de la Superintendencia de Industria y Comercio (SIC), se realiza una evaluación sistemática. El ARA/DPIA del framework cumple con estos requisitos, cubriendo aspectos como la creación de perfiles o el tratamiento a gran escala de datos sensibles.

Gestión de consentimiento. Cuando sea requerido, el consentimiento debe ser previo, informado, específico e inequívoco. Los mecanismos deben permitir su retiro con la misma facilidad con que se otorgó, registrarse de manera auditable y gestionar las consecuencias de su retirada.

Derechos de los titulares. Se establecen procedimientos operativos para garantizar los derechos de acceso, rectificación, supresión, oposición y portabilidad, siempre dentro de los marcos legales aplicables.

4.1.4.3. Seguridad y Resiliencia

Robustez técnica. Los controles aseguran un funcionamiento confiable mediante el manejo de errores y casos excepcionales, una degradación graciosa hacia comportamientos seguros, pruebas con datos adversarios y monitoreo de *drift* en el desempeño.

Ciberseguridad. Se implementan protecciones contra amenazas basadas en mejores prácticas como el NIST CSF o CIS Controls. Esto incluye defensas contra ataques específicos de *machine learning* (envenenamiento, evasión, extracción de modelos), validación robusta de entradas, cifrado de datos y gestión segura de credenciales.

Autenticación y procedencia de contenido (para GenAI). Cuando sea viable técnicamente, se incorporan marcas de agua o firmas digitales para detectar contenido sintético y mecanismos de procedencia que permitan rastrear su origen, reconociendo al mismo tiempo las limitaciones tecnológicas actuales.

Telemetría y respuesta a incidentes. Se implementa un *logging* integral, sistemas de alerta para eventos críticos y protocolos de respuesta con roles definidos, incluyendo la notificación a autoridades y titulares según lo dispuesto por la SIC en caso de brechas.

4.1.4.4. Transparencia y Explicabilidad

Divulgaciones ciudadanas. Se requiere una transparencia proactiva, notificando claramente cuando un ciudadano interactúe con un sistema de IA, informando sobre su propósito, el tipo de decisiones que toma y los derechos del ciudadano. Para sistemas de alto riesgo, se publica información adicional en portales institucionales.

Documentación técnica. Para garantizar la auditabilidad, se elaboran *Model Cards* (Mitchell et al., 2019) y *Data Sheets* (Gebru et al., 2018), documentando el modelo, los datos, la arquitectura y las decisiones de diseño, incluyendo el manejo de *trade-offs*.

Trazabilidad. Se mantiene una capacidad de auditoría mediante el *logging* de entradas, salidas y versiones, junto con la implementación de métodos de explicabilidad (SHAP, LIME) adecuados al modelo y al usuario, balanceando las necesidades de explicabilidad con el desempeño y la privacidad.

Inteligibilidad de explicaciones. Las explicaciones se adaptan al destinatario: simples y sin jerga para el ciudadano, técnicas para el operador y completas para el auditor.

4.1.4.5. Atención al Ciudadano

Canales de reclamación y apelación. Se disponen mecanismos accesibles, gratuitos y con plazos razonables para reportar problemas, apelar decisiones automatizadas y solicitar revisión humana, garantizando que las apelaciones sean resueltas por humanos con capacidad de decisión.

Integridad del servicio. Se monitorea la satisfacción ciudadana y se compara la calidad del servicio con la línea base pre-IA, detectando exclusiones involuntarias y asegurando alternativas de servicio.

Accesibilidad universal. El diseño cumple con los estándares WCAG 2.1 nivel AA, es compatible con tecnologías asistivas y ofrece contenido comprensible y alternativas para población con baja alfabetización digital.

4.1.4.6. Gestión de Terceros (Proveedores)

Debida diligencia en selección. Se aplica un checklist que evalúa la conformidad regulatoria, la madurez en IA responsable, la calidad de la documentación técnica, las capacidades de seguridad y el soporte para auditoría.

Cláusulas contractuales obligatorias. Los contratos incluyen Acuerdos de Procesamiento de Datos (DPA), claridad sobre propiedad intelectual, SLAs cuantificables, gestión de cambios, derecho a auditoría y cláusulas de continuidad y portabilidad.

Monitoreo de desempeño de proveedores. Se evalúa periódicamente el cumplimiento de los SLAs, se revisan incidentes y se toman decisiones informadas sobre la renovación o cambio de proveedor basadas en evidencia.

4.1.5. Capa 5: Métricas y KPI´s

La quinta capa del framework establece un sistema integral de métricas e indicadores clave de desempeño que permite medir, monitorear y demostrar el cumplimiento normativo, la gestión efectiva de riesgos, la calidad técnica y el impacto transformador de los sistemas de inteligencia artificial en el servicio público distrital. Esta capa responde directamente a la necesidad de medición basada en evidencia planteada por la normativa nacional (CONPES, 2024), que señala la necesidad de sistemas robustos de seguimiento y evaluación para asegurar que las políticas públicas de IA generen resultados cuantificables y sostenibles. En el contexto específico de la Secretaría Distrital de Gobierno de Bogotá, cuyo Plan Estratégico Institucional 2024-2028 establece objetivos ambiciosos de transformación digital y modernización administrativa, la implementación de un sistema de medición consistente con los principios del COBIT 2019 se convierte en un elemento diferenciador que permite demostrar el valor agregado de las inversiones en tecnología y la gobernanza efectiva de estos sistemas críticos para la administración pública distrital.

4.1.5.1. Marco Conceptual de Medición y Alineación con COBIT 2019

El sistema de métricas propuesto integra tres perspectivas complementarias de medición derivadas de los marcos internacionales de referencia, siendo especialmente importante la integración con el modelo COBIT 2019, que en sus procesos MEA (Monitor, Evaluate and Assess) establece que la medición de la conformidad y el desempeño de los procesos de tecnología de la información debe estar vinculada directamente con los objetivos corporativos y las metas de negocio. COBIT 2019 enfatiza que la medición no es un fin en sí mismo, sino un medio para verificar que los procesos de gobierno de TI generan valor y cumplen con los requisitos de control y conformidad establecidos. La perspectiva de cumplimiento regulatorio, alineada con el AI Act y CONPES 4144 (2024), incluye métricas que demuestran adherencia a requisitos normativos obligatorios, tales como la realización de evaluaciones de impacto ARA/DPIA (Análisis de Riesgos y Gestión de Datos Personales en Inteligencia Artificial), la documentación técnica mediante Model Cards y Data Sheets, y la implementación de controles específicos según el nivel de riesgo del sistema. Esta perspectiva operacionaliza el concepto de responsabilidad mediante indicadores verificables de cumplimiento que

pueden ser auditados tanto internamente como por entidades externas, cumpliendo así con los requisitos del COBIT 2019 en materia de monitoreo continuo de la conformidad regulatoria.

La perspectiva de gestión de riesgos, fundamentada en el NIST AI RMF e ISO/IEC 23894, incluye métricas que cuantifican la efectividad de los procesos de identificación, evaluación y mitigación de riesgos a lo largo del ciclo de vida de los sistemas de IA. Esta perspectiva es particularmente importante en el contexto de la Secretaría Distrital de Gobierno, dado que administra servicios públicos críticos cuyo mal funcionamiento o la discriminación podría afectar directamente a la ciudadanía bogotana. Los indicadores en esta perspectiva incluyen tasas de materialización de riesgos, tiempos de respuesta a incidentes identificados, y efectividad de controles implementados, permitiendo una gestión proactiva basada en evidencia cuantitativa. El COBIT 2019, en su proceso DSS (Deliver, Service and Support), establece que el monitoreo de incidentes y la capacidad de respuesta son elementos críticos para la entrega de servicios de TI de calidad, lo cual se adapta perfectamente al contexto de sistemas de IA cuya ejecución incorrecta podría acarrear consecuencias significativas.

La perspectiva de creación de valor público, desarrollada mediante un enfoque adaptado del Balanced Scorecard, incluye métricas que miden el impacto tangible de la IA en la eficiencia operativa, la calidad del servicio, la satisfacción ciudadana y la equidad en el acceso a los servicios. Esta perspectiva responde a la necesidad fundamental de que la adopción de IA en el sector público genere valor observable para la ciudadanía, no solo innovación tecnológica per se. El DAFP (Departamento Administrativo de la Función Pública) ha señalado en sus guías de gobierno de TI que la continuidad de las iniciativas de transformación digital depende críticamente de la capacidad de demostrar un impacto positivo en la calidad y la eficiencia del servicio público, lo cual se refleja en esta perspectiva de medición.

4.1.5.2. Principios de Diseño de las Métricas Vinculados a Niveles de Madurez COBIT 2019

Las métricas y KPIs del framework se diseñaron siguiendo principios específicos que aseguran su utilidad práctica y viabilidad operativa, especialmente considerando el contexto institucional particular de la Secretaría Distrital de Gobierno y sus capacidades actuales. El COBIT 2019 define seis niveles de capacidad para los procesos de gobierno de TI: Nivel 0 (Incomplete) donde los procesos no están implementados o no alcanzan los objetivos; Nivel 1 (Performed) donde los procesos se ejecutan informalmente; Nivel 2 (Managed) donde los procesos se planifican, ejecutan y monitorean; Nivel 3 (Defined) donde los procesos están estandarizados y documentados; Nivel 4 (Quantitatively Managed) donde los procesos son controlados cuantitativamente; y Nivel 5 (Optimizing) donde los procesos se mejoran continuamente. La proporcionalidad al riesgo es un principio fundamental que establece que

los requisitos de medición deben aumentar según el nivel de riesgo del sistema implementado. Los sistemas de alto riesgo, definidos como aquellos que afectan derechos fundamentales o beneficios significativos de la ciudadanía, requieren un monitoreo continuo automatizado que abarque múltiples dimensiones de equidad, privacidad y seguridad. Por el contrario, en sistemas de riesgo limited o mínimo se aplican métricas más simplificadas, con frecuencias de revisión menos exigentes, lo que permite optimizar la asignación de recursos de monitoreo. En términos de COBIT 2019, esto se alinea con el concepto de que distintos procesos y dominios pueden alcanzar distintos niveles de madurez según su criticidad para el negocio.

La viabilidad operativa es otro principio crítico que reconoce las capacidades heterogéneas de las entidades distritales y la necesidad de que el framework sea implementable de forma progresiva. Las métricas priorizan aquellas que pueden ser capturadas con infraestructura estándar, como logs de aplicaciones, encuestas de satisfacción, y registros administrativos, sobre aquellas que requieren herramientas especializadas costosas o procesos de recolección de datos complejos. La Secretaría Distrital de Gobierno, a través del Plan Estratégico de Tecnologías de la Información (PETI) vigente desde 2025, ha establecido una infraestructura básica de monitoreo que puede aprovecharse para la captura de estas métricas estándar. El principio de accionabilidad establece que cada métrica se conecta directamente con acciones de mejora o respuesta automática. Los umbrales de aceptabilidad definen cuándo se activan protocolos de revisión, mitigación o escalamiento, evitando así medición que no genere cambios o mejoras reales. Finalmente, la trazabilidad a objetivos estratégicos garantiza que las métricas de la Capa 5 se mapean directamente a las metas institucionales establecidas en la Capa 1 Principios y Políticas y a los objetivos del CONPES 4144 (2024), asegurando que todo lo que se mide contribuye realmente a lograr objetivos concretos.

4.1.5.3. Métricas de Cumplimiento Normativo y Ciclo de Vida Regulatorio

Este conjunto de métricas verifica que se cumplen los requisitos normativos del CONPES 4144 (2024), la Ley 1581 (2012) sobre protección de datos personales, las circulares de la Superintendencia de Industria y Comercio (SIC), y las reglas del AI Act aplicables por analogía a sistemas de alto riesgo en entidades públicas.

El porcentaje de casos de uso registrados en el sistema central constituye un indicador fundamental de trazabilidad institucional que mide la completitud del inventario distrital de sistemas de IA. La fórmula específica es: (Casos de uso registrados en el Registro Central / Total de casos de uso identificados en operación o desarrollo) × 100. El objetivo establecido es 100%, dado que la ausencia de registro imposibilita la aplicación sistemática de controles de gobernanza y representa un riesgo de cumplimiento normativo significativo. La frecuencia de medición es mensual, con auditoría trimestral

de completitud mediante cruce con registros de áreas de TI, contratación y presupuesto. El CONPES 4144 (2024) establece que la trazabilidad del ecosistema de IA es una condición necesaria para la gobernanza efectiva, mientras que la Ley 2279 (2022) exige transparencia en la implementación de tecnologías emergentes en el sector público. Un porcentaje inferior al 90% indica riesgos sistémicos de gobernanza y requiere acciones correctivas inmediatas del Comité de IA.

La cobertura de uso del Catálogo Distrital de Proveedores Pre-evaluados mide el grado de adopción de este instrumento de eficiencia en los procesos de contratación de soluciones de IA. La fórmula es: (Número de contrataciones que utilizaron proveedores del catálogo / Total de contrataciones de sistemas de IA durante el período) × 100. El objetivo diferenciado es ≥70% para sistemas de alto riesgo (donde la evaluación previa mitiga riesgos críticos) y ≥50% para sistemas de riesgo limitado (permitiendo mayor flexibilidad). La frecuencia es trimestral, con consolidación anual para análisis de tendencias. La Ley 2279 (2022) promueve la eficiencia en la contratación pública mediante la estandarización de procesos y la reducción de cargas administrativas. El uso sistemático del catálogo permite reducir entre 40-60% el tiempo de evaluación técnica en licitaciones, fortalece el poder de negociación institucional mediante compras agregadas y facilita la transferencia de lecciones aprendidas entre entidades. Un porcentaje de adopción inferior al 40% sugiere barreras operativas que deben ser investigadas y removidas.

El porcentaje de casos de uso con clasificación de riesgo documentada es una métrica fundamental que proporciona visibilidad sobre qué tantos sistemas han sido clasificados según el marco de riesgos. La fórmula específica es: (Casos de uso con clasificación de riesgo formalizada / Total de casos de uso de IA en operación o desarrollo) × 100. El objetivo establecido es 100%, ya que clasificar el riesgo es el primer paso obligatorio del framework y sin ello no es posible aplicar controles apropiados. La frecuencia de medición es trimestral, con revisión completa anual para identificar casos nuevos. El CONPES 4144 (2024) establece que la gestión basada en riesgos debe ser el eje transversal de la gobernanza de IA en Colombia, y el AI Act de la Unión Europea proporciona un modelo que el Gobierno Nacional ha adoptado como referencia conceptual para dividir sistemas en cuatro niveles de riesgo.

El porcentaje de sistemas de alto riesgo con ARA/DPIA completo y aprobado es otra métrica crítica que verifica que los sistemas que manejan datos personales hayan sido sometidos a un análisis riguroso de riesgos. La fórmula es: (Sistemas de alto riesgo con ARA/DPIA realizado según plantilla del toolkit / Total de sistemas de alto riesgo) × 100. El objetivo es 100%, indicando que ningún sistema de alto riesgo puede ponerse en funcionamiento sin que el ARA/DPIA haya sido aprobado por el Comité de IA. La frecuencia es verificación continua antes de cada despliegue, consolidado trimestralmente en reportes de gobernanza. La Circular Externa 002-2024 de la SIC establece que es obligatorio realizar

Análisis de Riesgos y Gestión (ARA) para tratamientos de datos personales con riesgos significativos, lo que aplica directamente a sistemas de IA que procesan datos personales.

El porcentaje de sistemas con documentación técnica completa (Model Cards y Data Sheets) mide la capacidad de transparencia y auditabilidad del ecosistema de IA. La fórmula es: (Sistemas con Model Card y Data Sheets actualizados / Total de sistemas en operación) × 100. Los objetivos diferenciados son: 100% para sistemas de alto riesgo (donde la documentación es indispensable para propósitos de auditoría regulatoria), 80% para riesgo limitado en el primer año progresando a 100% en años subsecuentes. La frecuencia es semestral, reconociendo que los sistemas evolucionan y requieren la actualización de su documentación. El CONPES 4144 (2024) establece que los sistemas de IA deben disponer de información técnica accesible para auditores, supervisores, y cuando sea apropiado, para los usuarios finales.

El porcentaje de reuniones del Comité de IA realizadas según calendario es una métrica de gobernanza que verifica que la estructura de gobernanza está funcionando de verdad. La fórmula es: (Reuniones del Comité de IA efectivamente realizadas / Reuniones programadas según política) × 100. El objetivo es ≥ 90%, reconociendo que circunstancias excepcionales pueden requerir reprogramación pero sin permitir que esto se convierte en práctica regular. La frecuencia es trimestral, consolidando los registros de asistencia y actas.

La cobertura de capacitación obligatoria en IA Responsable y Generativa es una métrica compuesta que evalúa si se está desarrollando la capacidad institucional necesaria para la adopción responsable de estas tecnologías. La fórmula específica es: (Funcionarios certificados vigentemente en IA Responsable y/o IA Generativa según rol / Total de funcionarios en roles clave + usuarios activos de herramientas GenAI) × 100. Los roles considerados incluyen: (a) roles de gobernanza: miembros del Comité de IA, Oficial de IA Responsable, DPO, auditores internos; (b) roles técnicos: responsables técnicos de casos de uso, desarrolladores, arquitectos de datos; (c) roles operativos: sponsors de negocio, funcionarios con acceso a herramientas de IA Generativa institucionales.

Los objetivos se establecen de forma diferenciada según criticidad: 100% para roles de gobernanza y técnicos (capacitación obligatoria previa al ejercicio del rol), 100% para usuarios de IA Generativa antes de habilitar acceso a herramientas (requisito de seguridad no negociable), y ≥80% para roles de soporte (meta progresiva). La frecuencia de medición es trimestral, con verificación continua integrada a los sistemas de autenticación de herramientas GenAI que validan certificación vigente antes de conceder acceso.

La certificación en IA Responsable tiene vigencia de 12 meses con actualización anual de contenidos, mientras que la certificación en IA Generativa tiene vigencia de 12 meses pero requiere actualización

trimestral de contenidos dada la velocidad de evolución tecnológica y normativa en este dominio específico. La recertificación anticipada es obligatoria ante cambios normativos significativos o identificación de brechas de conocimiento mediante análisis de incidentes.

El CONPES 4144 (2024) establece explícitamente que el desarrollo de capacidades es una condición habilitadora crítica para que el sector público pueda adoptar IA de manera responsable, reconociendo que la tecnología sin capital humano preparado genera riesgos inaceptables. La Ley 2279 (2022) de Transformación Digital establece que las entidades públicas deben garantizar la alfabetización digital de sus funcionarios en tecnologías emergentes como condición para su implementación efectiva. Un porcentaje de cobertura inferior al 70% en roles críticos representa un riesgo operativo severo que debe activar protocolos de mitigación inmediata, incluyendo la posible suspensión temporal de acceso a sistemas críticos hasta completar la capacitación requerida.

4.1.5.4. Métricas de Gestión de Riesgos y Monitoreo de Incidentes

Estas métricas hacen funcional el proceso de gestión de riesgos establecido en la Capa 1 Política de Gestión de Riesgos y permiten verificar cuantitativamente si los controles implementados en la Capa 4 están funcionando. El número de incidentes de IA por tipología proporciona visibilidad clara sobre qué tipos de problemas experimentan los sistemas implementados. Las categorías contempladas incluyen incidentes técnicos (fallos operacionales, errores de predicción, indisponibilidad del servicio), incidentes de privacidad (brechas de datos, accesos indebidos a información personal), incidentes de equidad (discriminación detectada en decisiones automatizadas, sesgos que generan resultados discriminatorios), incidentes de seguridad (ataques adversarios contra modelos, vulnerabilidades explotadas), y quejas ciudadanas (retroalimentación de usuarios sobre problemas no formalmente categorizados). El objetivo es una tendencia decreciente trimestre a trimestre, con una tasa de incidentes críticos por debajo del umbral definido según nivel de riesgo del sistema específico. La frecuencia es registro continuo con reporte consolidado mensual. El proceso DSS01 (Manage Operations) de COBIT 2019 establece que las organizaciones deben monitorear continuamente el desempeño operativo de sus sistemas, identificar problemas y tomar acciones correctivas de manera oportuna.

El tiempo medio de respuesta a incidentes es una métrica que evalúa la capacidad de la organización de reaccionar ante problemas identificados. La fórmula específica es: Σ (tiempo desde detección hasta resolución completa) / Número total de incidentes procesados. Los objetivos diferenciados según criticidad son: ≤ 24 horas para incidentes críticos (aquellos que afectan derechos fundamentales o generan exclusión de ciudadanos), ≤ 72 horas para incidentes de alto impacto, ≤ 1 semana para incidentes de impacto medio. La frecuencia es consolidación mensual con alertas automatizadas cuando se violan los tiempos máximos de respuesta. La tasa de materialización de riesgos mide si el análisis de riesgos fue suficientemente profundo. La fórmula es: (Riesgos que efectivamente se presentaron durante período / Total de riesgos identificados en ARA/DPIA durante período anterior) × 100. El objetivo es < 5%, lo que indicaría que el análisis de riesgos fue suficientemente completo y que los controles preventivos fueron efectivos en prevenir la mayoría de riesgos identificados. Una tasa baja valida la calidad del proceso de análisis de riesgos y demuestra que las mitigaciones están funcionando. La frecuencia es semestral permitiendo acumular suficientes incidentes para análisis estadístico.

La efectividad de controles es una métrica cualitativa pero fundamentada en evidencia que evalúa si los controles implementados están funcionando como fueron diseñados. La evaluación se realiza durante auditorías internas por el equipo de control interno, o externas por auditores independientes

certificados. El objetivo es ≥ 85% de los controles evaluados como efectivos, lo que indica que la mayoría de los mecanismos de protección están operando correctamente. La frecuencia es anual mediante auditoría interna para todos los sistemas, y bianual mediante auditoría externa independiente específicamente para sistemas de alto riesgo. Los procesos MEA (Monitor, Evaluate and Assess) del COBIT 2019 establecen que las organizaciones deben evaluar periódicamente si sus controles continúan cumpliendo objetivos.

4.1.5.5. Métricas de Calidad de Datos y Confiabilidad de Fuentes

Reconociendo que la calidad de los datos determina de forma crítica la calidad, equidad y confiabilidad de los sistemas de IA, estas métricas evalúan las dimensiones fundamentales del ciclo de vida de datos desde su recolección hasta su depuración y uso en modelos. La completitud de datos mide qué porcentaje de registros contienen valores válidos en campos críticos. La fórmula es: (Registros sin valores faltantes en campos críticos / Total de registros en dataset) × 100. El objetivo es ≥ 95% para campos críticos, reconociendo que un porcentaje bajo de datos faltantes puede introducir sesgos si la ausencia no es aleatoria. La frecuencia es evaluación mensual para datasets en uso activo, facilitando detección temprana de problemas de calidad. El CONPES 4144 (2024) enfatiza explícitamente que contar con datos de buena calidad es un pilar fundamental de la estrategia nacional de IA, reconociendo que datos deficientes limitan significativamente los beneficios potenciales.

La tasa de detección de drift (cambio de datos o cambio de patrón en el tiempo) mide la capacidad del sistema de identificar cambios en la distribución de datos o en la relación entre variables a lo largo del tiempo. Esta es una métrica importante porque el drift es una causa frecuente de que los modelos pierdan desempeño en producción sin que se haya anticipado durante la fase de entrenamiento. La métrica es: Número de casos donde se detectó drift significativo durante monitoreo. El objetivo es detección temprana antes de que el desempeño del modelo se deteriore de forma severa. La frecuencia depende de la criticidad del sistema: semanal para sistemas de alto riesgo, mensual para riesgo limitado. La representatividad demográfica es una métrica que evalúa si la distribución de datos refleja adecuadamente la población objetivo en dimensiones protegidas como género, edad, estrato socioeconómico, y localidad geográfica. La evaluación se realiza mediante tests estadísticos como Chi-cuadrado o pruebas de similitud distribucional. El objetivo es un p-valor ≥ 0.05 indicando que no hay diferencia estadísticamente significativa entre la distribución de datos y la población objetivo, o justificación documentada y aprobada de diferencias intencionales. La frecuencia es trimestral o con cada actualización significativa de datos de entrenamiento. El índice de calidad de documentación de datos evalúa si la documentación sobre los datos es suficientemente completa en dimensiones clave como origen de datos, composición del dataset, procesos de limpieza aplicados, sesgos identificados,

y limitaciones conocidas. La evaluación se realiza en una escala 0-100 basada en un checklist integral. El objetivo es ≥ 80%. La frecuencia es con cada nuevo conjunto de datos o actualización importante. La documentación rigurosa de datos es un requisito de transparencia y auditabilidad establecido en los marcos internacionales.

4.1.5.6. Métricas de Desempeño Técnico y Confiabilidad de Modelos

Estas métricas evalúan el rendimiento técnico de los sistemas de IA, asegurando que cumplan con estándares de precisión, equidad, disponibilidad y velocidad de respuesta apropiados para el contexto del servicio público distrital. Las métricas de precisión varían según el tipo de sistema: para sistemas de clasificación se utilizan Precisión (qué porcentaje de predicciones positivas fueron correctas), Recall o Sensibilidad (qué porcentaje de casos positivos fueron identificados correctamente), F1-Score (promedio ponderado de Precisión y Recall), y AUC-ROC (medida de qué tan bien el modelo diferencia entre categorías). Para sistemas de regresión se utilizan MAE (Error Absoluto Medio), RMSE (Raíz del Error Cuadrático Medio), y R^2 (qué proporción de variación es explicada por el modelo). Para sistemas de generación de texto se utilizan BLEU y ROUGE (medidas de similitud de texto) además de evaluación humana de calidad. El objetivo de precisión es un umbral específico definido durante la fase de análisis de riesgos en ARA/DPIA, típicamente ≥ 85% para casos críticos donde errores tienen consecuencias significativas. La frecuencia es monitoreo continuo con alertas automatizadas cuando se cruzan umbrales de alerta, con reporte consolidado semanal o mensual. El desempeño técnico insuficiente compromete directamente la confiabilidad del servicio público y puede generar consecuencias adversas para ciudadanos.

Las métricas de equidad (fairness) cuantifican el grado de discriminación potencial del sistema. El Disparate Impact Ratio es: (Tasa de resultado positivo en grupo protegido) / (Tasa de resultado positivo en grupo mayoritario). El Equal Opportunity Difference es la diferencia absoluta en tasas de acierto entre grupos demográficos. El objetivo es un Disparate Impact ≥ 0.8 (indicando que el grupo protegido recibe resultados positivos en al menos 80% de la tasa del grupo mayoritario) y Equal Opportunity Difference ≤ 0.1 (diferencia menor a 10 puntos porcentuales, ajustable según contexto específico). La frecuencia es semanal o mensual en monitoreo continuo. El AI Act y el CONPES 4144 (2024) establecen explícitamente que la no discriminación es un principio fundamental que debe permear todos los sistemas de IA del sector público.

La disponibilidad del sistema (uptime o tiempo en que el servicio está activo) es una métrica operacional crítica en el contexto de servicios públicos. La fórmula es: (Tiempo de operación efectiva / Tiempo total planificado) × 100. El objetivo es ≥ 99% para servicios críticos cuya indisponibilidad afectaría derechos de ciudadanos, y ≥ 95% para servicios no críticos. La frecuencia es monitoreo

continuo por infraestructura de TI con reporte consolidado mensual. El tiempo de respuesta mide cuán rápido el sistema responde a las solicitudes de usuarios. La métrica específica es el Percentil 95 del tiempo de respuesta a consultas o inferencias del modelo. El objetivo es un umbral definido en acuerdos de nivel de servicio (SLA), típicamente ≤ 2 segundos para interacciones ciudadanas directas en portales web o aplicaciones móviles. La frecuencia es monitoreo continuo con reporte semanal. La experiencia de usuario determina críticamente si la población adopta y está satisfecha con servicios digitales.

4.1.5.7. Métricas de Impacto en Servicio Público y Satisfacción Ciudadana

Este conjunto de métricas evalúa el valor real que generan los sistemas de IA, respondiendo a la pregunta fundamental: Está la IA mejorando efectivamente los servicios públicos y la experiencia ciudadana de manera medible?. La satisfacción ciudadana se mide mediante encuestas post-interacción que capturan el CSAT (Customer Satisfaction Score) medido en escala 1-5, y NPS (Net Promoter Score) calculado como: (% de personas que recomendarían - % de personas que no recomendarían). El objetivo es ≥ 80% de respondientes satisfecho o muy satisfecho, y NPS ≥ 50 indicando que la mayoría de ciudadanos recomendaría el servicio. La frecuencia es captura continua mediante muestreo estadístico con consolidación mensual. La aceptación ciudadana de la IA en el sector público depende fundamentalmente de si la población experimenta mejoras reales en los servicios que recibe.

La eficiencia operativa es una métrica que compara cómo funcionaban las cosas antes y después de implementar IA. Las métricas específicas son: (1) Tiempo promedio de atención o resolución de trámite, usando los datos previos a IA como comparación, y (2) Reducción de costos operativos. El objetivo es mejora ≥ 20% respecto a cómo era antes, un umbral conservador que reconoce que no todos los casos de uso generarán ganancias de eficiencia equivalentes. La frecuencia es evaluación trimestral. El CONPES 4144 (2024) establece explícitamente que eficiencia y productividad son objetivos clave de la adopción de IA en el sector público.

El volumen de servicio mide la capacidad de atender más transacciones con la implementación de IA. La métrica es: Número de transacciones o consultas procesadas por el sistema de IA. Una métrica complementaria es la tasa de resolución en primer contacto: (Casos resueltos sin escalamiento a humano / Total de casos procesados) × 100. El objetivo es crecimiento sostenido del volumen de transacciones, con tasa de resolución ≥ 70% indicando que el sistema es capaz de manejar la mayoría de las solicitudes de ciudadanos. La frecuencia es consolidación mensual. La escalabilidad es un beneficio clave esperado de la automatización con IA, permitiendo al sector público servir a más ciudadanos con igual o menor inversión presupuestal.

La tasa de escalamiento a humano mide el balance entre dejar que el sistema trabaje solo y el nivel de supervisión que se necesita. La fórmula es: (Casos escalados a revisión o decisión humana / Total de casos procesados) × 100. El objetivo es un balance pragmático, típicamente 5-15% según criticidad del servicio, reconociendo que un porcentaje muy alto indica que el sistema no es suficientemente independiente para generar eficiencia, mientras que un porcentaje muy bajo puede indicar que no se está revisando suficientemente y hay riesgos de problemas no detectados. La frecuencia es consolidación semanal o mensual. La equidad de acceso es una métrica que evalúa si todos los segmentos de la población bogotana están accediendo equitativamente al servicio. Se realiza análisis desagregado de uso por grupos demográficos (género, edad, estrato, localidad), para identificar si hay exclusión o subrepresentación de grupos vulnerables. El objetivo es no diferencias estadísticamente significativas entre grupos, sin justificación por diferencias en elegibilidad. La frecuencia es evaluación trimestral. El principio constitucional de igualdad exige que los servicios públicos no generen exclusión digital ni discriminación por proximidad geográfica o estrato socioeconómico.

4.1.5.8. Métricas de Madurez Institucional y Sostenibilidad

Este conjunto de métricas evalúa qué tan desarrolladas están las capacidades de gobernanza de IA dentro de la Secretaría, facilitando la planificación estratégica y la asignación de recursos a lo largo del tiempo. El nivel de madurez institucional se evalúa según un modelo de madurez 0-3 que será descrito en detalle en la sección 4.1.6 del framework. Brevemente, Nivel 0 indica iniciativas incipientes sin estructura formal, Nivel 1 indica procesos informales y inconsistentes, Nivel 2 indica procesos documentados y parcialmente implementados, y Nivel 3 indica procesos plenamente implementados, monitoreados y sujetos a mejora continua. El objetivo es avanzar de al menos un nivel cada año, reconociendo que el desarrollo institucional es un proceso gradual. La frecuencia es evaluación anual mediante autoevaluación del Comité de IA con validación externa por consultores especializados.

La métrica de cobertura de capacitación obligatoria, detallada en la subsección 4.1.5.3 como componente de cumplimiento normativo, constituye simultáneamente un indicador de madurez institucional. Las organizaciones en Nivel 0 (Incipiente) típicamente presentan cobertura inferior al 30% y sin sistematización; en Nivel 1 (Informal) la cobertura oscila entre 30-60% con contenidos heterogéneos; en Nivel 2 (Documentado) la cobertura supera el 80% en roles clave con programas estructurados; y en Nivel 3 (Optimizado) se alcanza 100% en roles críticos con actualización continua y evaluación de efectividad mediante análisis de reducción de incidentes atribuibles a brecha de conocimiento.

El porcentaje de trámites y servicios con IA implementada es una métrica de transformación digital que mide cuántos de los servicios que ofrece la Secretaría ya tienen componentes de IA. La fórmula

es: (Servicios con componentes de IA / Total de servicios candidatos identificados en roadmap) × 100. El objetivo es crecimiento progresivo según roadmap aprobado, por ejemplo: 20% en el año 1, 40% en el año 2, 60% en el año 3, reconociendo que la transformación requiere tiempo y debe hacerse de forma ordenada. La frecuencia es evaluación semestral. La relación costo-beneficio es una métrica de sostenibilidad financiera que evalúa si la inversión en IA está generando retorno. La fórmula es: (Beneficios tangibles monetarios + valor estimado de beneficios intangibles) / (Costos totales de implementación incluyendo desarrollo, infraestructura, operación y soporte) sobre período de 3 años. El objetivo es ROI ≥ 1.5 a 3, indicando que por cada peso invertido se generan entre 1.50 y 3 pesos de valor. La frecuencia es evaluación anual. La sostenibilidad fiscal exige demostración clara de que la inversión pública en tecnología está generando valor.

El índice de sostenibilidad institucional incorpora como dimensión específica la "capacidad de autosuficiencia en capacitación", evaluando si la entidad depende críticamente de consultores externos para formación o si ha desarrollado capacidades internas de formadores certificados. Las entidades con puntuación ≥70 en sostenibilidad típicamente cuentan con al menos 2-3 formadores internos certificados por cada 100 funcionarios usuarios de IA, permitiendo escalabilidad y actualización ágil de contenidos sin dependencia presupuestal externa.

4.1.5.9. Dashboard Integrado y Reportería

Las métricas descritas se integran en un Dashboard de Gobernanza de IA que proporciona visibilidad clara del estado de los sistemas de IA. El dashboard se estructura en tres niveles de información: Vista ejecutiva para Alta Dirección que muestra indicadores resumen por dimensión (cumplimiento, riesgos, impacto), con colores verde/amarillo/rojo según umbrales, actualizado mensualmente para facilitar toma de decisiones. Vista operativa para gestores de sistemas que muestra métricas técnicas y de servicio para cada caso de uso específico, con alertas automatizadas cuando se superan umbrales, actualización continua para permitir respuesta rápida a problemas. Vista de auditoría para control interno y auditores que muestra trazabilidad completa de cumplimiento normativo, evidencias de que los controles están funcionando, registro de incidentes, actualización trimestral. Este sistema de medición transforma la gobernanza de IA de una aspiración normativa en una práctica basada en evidencia cuantitativa observable, permitiendo la mejora continua, la rendición de cuentas clara, y la demostración concreta de valor público generado por la adopción responsable de inteligencia artificial en el Distrito Capital de Bogotá. Los componentes del Dashboard se referencian en el Anexo C.

4.1.6. Modelo de Madurez Institucional

El modelo de madurez institucional proporciona un marco de referencia para evaluar y orientar el progreso de la Secretaría Distrital de Gobierno en la adopción responsable de la inteligencia artificial.

Este modelo se fundamenta en los principios de COBIT 2019, que establece una escala de madurez de cinco niveles para los procesos de gobierno de TI, adaptándose específicamente al contexto de gobernanza de IA en entidades públicas del Distrito Capital. El modelo propuesto contempla cuatro niveles de madurez (Nivel 0 a Nivel 3) reconociendo las capacidades actuales de las entidades distritales y estableciendo una trayectoria clara de evolución institucional que responde a los objetivos del CONPES 4144 (2024) de desarrollar capacidades progresivas en el ecosistema de IA colombiano.

Nivel 0: Incipiente representa la situación inicial donde no existe gobernanza formal de IA. A este nivel, las iniciativas de IA son aisladas, informales y carecen de coordinación institucional. No hay políticas documentadas, ni procesos establecidos para la gestión de riesgos, y la decisión de adoptar IA se toma de forma reactiva respecto a necesidades operativas puntuales sin evaluación integral de impacto. La infraestructura de monitoreo no existe, o es mínima. Las capacidades técnicas se concentran en individuos específicos sin transferencia de conocimiento. Los requisitos de conformidad normativa no se abordan de forma sistemática. La Secretaría podría estar en este nivel si inicia exploraciones puntuales de IA sin estructura institucional, identificable cuando solo algunos departamentos experimentan con soluciones de IA sin coordinación central.

Nivel 1: Informal marca el primer paso donde comienzan a surgir iniciativas coordinadas de IA pero sin formalización completa. Se establece un Comité de IA incipiente que se reúne de forma irregular. Se inician procesos de evaluación de riesgos pero de manera inconsistente, aplicados solo a algunos sistemas. Existe documentación parcial de políticas y procedimientos, pero estos no se comunican ni se cumplen sistemáticamente. Se realizan algunas capacitaciones en IA responsable para ciertos grupos, pero sin cobertura integral. Los controles son parcialmente implementados y se detectan incidentes pero la respuesta es reactiva. Este nivel refleja una transición donde la organización reconoce la necesidad de gobernanza pero aún opera con muchos aspectos informales y personas clave.

Nivel 2: Documentado representa un punto de madurez intermedia donde la gobernanza de IA está parcialmente implementada y documentada. A este nivel, el Comité de IA funciona regularmente con actas y trazabilidad de decisiones. Se cuenta con políticas formales de gobernanza de IA comunicadas a toda la organización, aunque su cumplimiento requiere supervisión activa. Los procesos de gestión de riesgos están documentados y aplicados a todos los sistemas de alto riesgo, aunque con variabilidad en profundidad y calidad. La documentación técnica de sistemas existe aunque puede presentar gaps. Se realiza capacitación sistemática en roles clave, con cobertura de ≥80% en roles de gobernanza. Los controles están implementados para sistemas críticos, monitoreados de forma regular. Los incidentes

se registran, clasifican y se inician protocolos de respuesta. La infraestructura de monitoreo básica está en operación. Este nivel es alcanzable en el mediano plazo con inversión en formalización de procesos.

Nivel 3: Optimizado representa la madurez institucional objetivo a tres años, donde la gobernanza de IA está completamente implementada, documentada, monitoreada y sometida a mejora continua. A este nivel, el Comité de IA funciona operativamente con alta efectividad, tomando decisiones informadas con datos. Las políticas están plenamente implementadas con cumplimiento ≥90%. Los procesos de gestión de riesgos son robustos, aplicados consistentemente a todos los sistemas con documentación completa de ARA/DPIA y controles. La documentación técnica es exhaustiva, actualizada y accesible. La capacitación alcanza 100% en roles clave con actualizaciones continuas. Los controles son altamente efectivos (≥85%), monitoreados en tiempo real. La respuesta a incidentes es ágil y efectiva. El dashboard de gobernanza proporciona visibilidad en tiempo real de todas las dimensiones. La cultura organizacional valora la IA responsable. Se genera valor público cuantificable. La sostenibilidad institucional es alta con apropiación de líderes, continuidad presupuestal y capacidades internas robustas.

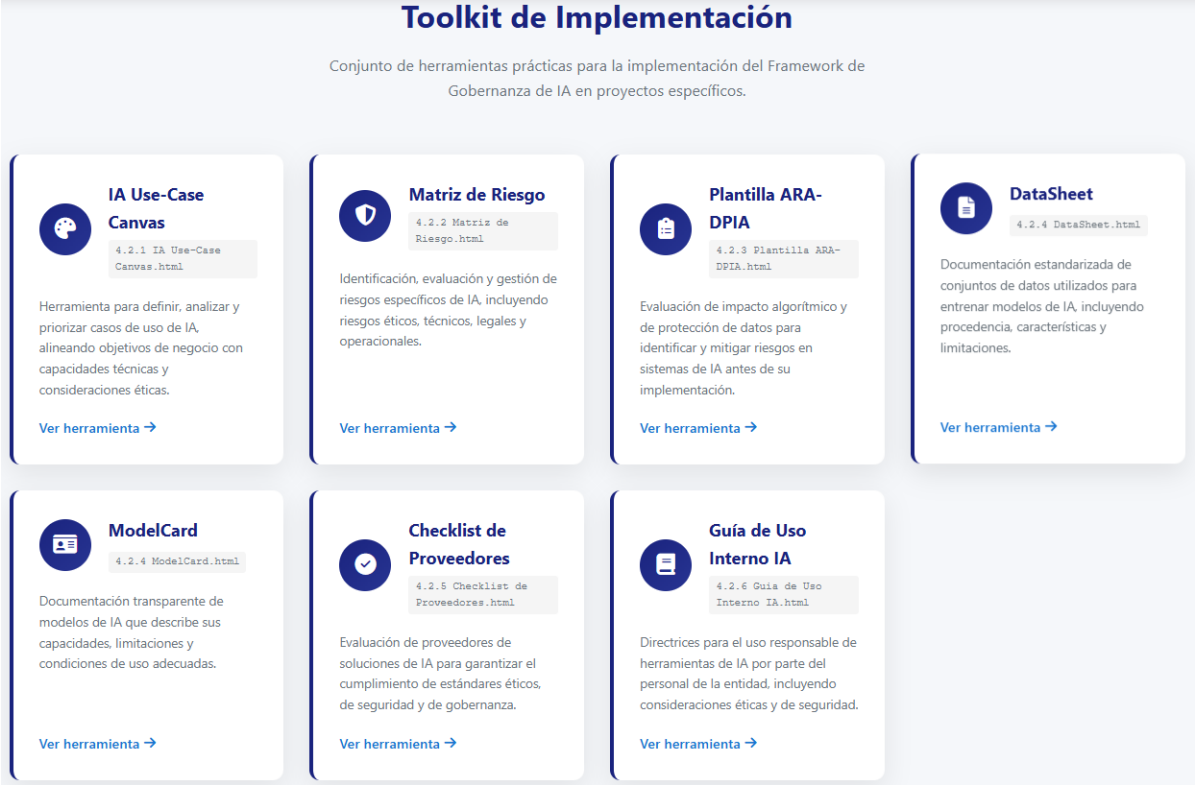
El movimiento entre niveles debe ser progresivo, con evaluaciones anuales formales mediante autoevaluación del Comité de IA validada por auditores externos independientes. Cada nivel incorpora las capacidades del anterior, generando una progresión que es realista pero ambiciosa. El CONPES 4144 (2024) enfatiza que la madurez institucional es determinante para la adopción responsable y sostenible de IA. El modelo propuesto permite a la Secretaría Distrital de Gobierno comunicar claramente dónde se encuentra, qué capacidades necesita desarrollar, y cuáles son los hitos de progreso esperados, facilitando la planificación estratégica y la asignación de recursos para convertir la gobernanza de IA en una capacidad institucional permanente y efectiva que genere confianza pública y valor sostenible.

4.2. Caja de Herramientas Operativa (Toolkit)

El toolkit representa el componente práctico del framework, al ofrecer un conjunto de plantillas, matrices y guías listas para su aplicación, cuyo propósito es facilitar la implementación de la gobernanza de la inteligencia artificial en entidades que presentan niveles diversos de capacidad técnica. Cada uno de los instrumentos ha sido concebido con criterios de usabilidad, mediante el uso de lenguaje claro, estructuras intuitivas e instrucciones detalladas paso a paso; de adaptabilidad, al incluir campos obligatorios y opcionales claramente diferenciados, así como escalabilidad conforme al nivel de riesgo; de trazabilidad, al incorporar referencias explícitas a requisitos normativos como el CONPES 4144 (2024), la Ley 1581 (2012), el AI Act y los estándares ISO; y de integrabilidad, al asegurar su compatibilidad tanto entre sí como con el proceso de ciclo de vida definido por el framework.

El toolkit representa el componente práctico del framework, al ofrecer un conjunto de plantillas, matrices y guías listas para su aplicación, cuyo propósito es facilitar la implementación de la gobernanza de la inteligencia artificial en entidades que presentan niveles diversos de capacidad técnica. En la Figura 4 se presenta una visión general y la interrelación de los siete instrumentos que componen este toolkit operativo, los cuales se detallarán a continuación.

Figura 4 Toolkit de Implementación



Fuente: Elaboración propia (2025).

4.2.1. AI Use-Case Canvas

El propósito de esta herramienta consiste en estructurar y documentar de manera sistemática la propuesta de caso de uso durante la fase inicial de intake. Su objetivo es garantizar una comprensión clara del propósito del sistema, una identificación precisa de los actores involucrados, así como una evaluación preliminar de los riesgos asociados y de la viabilidad técnica y legal del proyecto. La Figura 5 ilustra la estructura completa y las doce secciones que componen el AI Use-Case Canvas, herramienta fundamental para el proceso de gobernanza.

La plantilla correspondiente se organiza en doce secciones, las cuales se describen detalladamente a continuación.

El manual de usuario correspondiente a esta herramienta se encuentra en el Anexo F1.

Sección 1: Identificación del Caso de Uso – Se incluye un título descriptivo que sintetiza la finalidad del sistema propuesto, junto con la identificación de la entidad responsable y su unidad organizacional correspondiente. Asimismo, se requiere consignar los datos del sponsor de negocio, incluyendo su nombre, cargo y medios de contacto, así como los del responsable técnico, especificando igualmente su nombre, cargo y datos de contacto. Esta sección concluye con la indicación de la fecha de elaboración del documento y la versión correspondiente, lo que permite establecer un control de versiones y asegurar la trazabilidad del proceso.

Sección 2: Contexto y Propósito - Describe de forma concisa el problema o necesidad que se pretende abordar, limitando la extensión a un máximo de 200 palabras para garantizar claridad y enfoque. En este apartado se debe especificar el servicio, trámite o proceso institucional al que se aplicará la solución basada en inteligencia artificial, así como el propósito específico del sistema, es decir, la función que se espera que cumpla dentro del flujo operativo. También se debe indicar el tipo de sistema de IA que se propone implementar, ya sea de clasificación, regresión, generación de lenguaje, recomendación u otro, según corresponda. Finalmente, se deben detallar los resultados esperados, procurando que estos sean cuantificables en la medida de lo posible, con el fin de facilitar su posterior evaluación y seguimiento.

Sección 3: Actores Involucrados - La identificación de los actores involucrados en el sistema contempla, en primer lugar, a los usuarios finales, entre los que se incluyen funcionarios públicos, ciudadanos y otros grupos que interactúan directamente con la solución tecnológica. El perfil de estos usuarios se caracteriza por niveles diversos de alfabetización digital, condiciones particulares de accesibilidad y una amplia diversidad sociocultural, lo que exige enfoques inclusivos en el diseño y despliegue del sistema. El responsable de los datos utilizados, identificado como el propietario de los datos, asume la obligación de garantizar su integridad, legalidad y uso ético. Además, se reconocen stakeholders que podrían verse afectados por las decisiones automatizadas del sistema, lo que implica una evaluación anticipada de impactos. En cuanto a la gobernanza, se definen roles específicos que incluyen al Delegado de Protección de Datos (DPO), las áreas de Tecnologías de la Información y Seguridad, el equipo jurídico y el Comité de Inteligencia Artificial, todos ellos con funciones diferenciadas en la supervisión, validación y toma de decisiones estratégicas.

Sección 4: Datos Requeridos - El funcionamiento del sistema depende de un conjunto de datos provenientes de fuentes internas institucionales, externas gubernamentales y de terceros autorizados.

Estas fuentes aportan categorías de datos que incluyen información personal, datos sensibles y registros públicos, cuya combinación permite alimentar los modelos de análisis y toma de decisiones. El volumen estimado de datos, así como la frecuencia de actualización, se determinan en función de la naturaleza del servicio y de los requerimientos operativos. Una evaluación preliminar de la calidad y representatividad de los datos permite identificar posibles sesgos, vacíos informativos o inconsistencias que podrían afectar el desempeño del sistema. Asimismo, se realiza una evaluación inicial en materia de privacidad, en la que se determina si el conjunto de datos incluye información personal o sensible, lo que activa protocolos específicos de protección conforme a la legislación vigente y a los estándares internacionales de gobernanza de datos.

Sección 5: Revisión Legal y Bases Jurídicas - La revisión legal del sistema parte de la identificación de la base normativa que habilita la prestación del servicio público correspondiente, considerando tanto disposiciones generales como específicas del sector involucrado. En los casos en que se contempla el tratamiento de datos personales, se analizan las bases jurídicas aplicables, entre las que se incluyen el consentimiento informado, la ejecución de contratos, el cumplimiento de obligaciones legales, el interés legítimo y el ejercicio de funciones públicas. Este análisis se complementa con la identificación de normativa específica aplicable, ya sea de carácter sectorial o territorial, que pueda incidir en el diseño, operación o supervisión del sistema. Finalmente, el área Jurídica realiza una evaluación preliminar de conformidad legal, con el propósito de anticipar ajustes normativos necesarios y asegurar la alineación del sistema con el marco regulatorio vigente.

Sección 6: Clasificación Preliminar de Riesgo - La clasificación inicial del riesgo se realiza conforme a los criterios establecidos en el Reglamento de Inteligencia Artificial (AI Act), los cuales permiten determinar si el sistema afecta derechos fundamentales, interviene en servicios esenciales, participa en decisiones que impactan directamente a personas o incorpora tecnologías de reconocimiento biométrico. A partir de esta evaluación, se asigna una clasificación preliminar que puede ubicarse en las categorías de riesgo inaceptable, alto, limitado o mínimo. Esta categorización se sustenta en una justificación técnica y normativa que considera el contexto de uso, el tipo de datos procesados, el grado de automatización y el impacto potencial sobre los derechos de los individuos, así como sobre la operación institucional.

Sección 7: Identificación Preliminar de Riesgos - La evaluación inicial de riesgos contempla dimensiones éticas, de privacidad, seguridad, equidad, operación y reputación institucional. En el plano ético, se identifican posibles afectaciones a la autonomía individual, la dignidad humana y la aparición de sesgos discriminatorios. En cuanto a la privacidad, se reconocen riesgos asociados con la re-identificación de datos, el uso secundario no autorizado y la existencia de brechas de protección.

Los riesgos de seguridad incluyen vulnerabilidades frente a ataques cibernéticos, fallos sistémicos y posibles manipulaciones maliciosas. Desde una perspectiva de equidad, se advierte la posibilidad de sesgos algorítmicos que generen exclusión de grupos poblacionales específicos. En el ámbito operativo, se consideran factores como la dependencia de proveedores externos, la sostenibilidad técnica del sistema y su eventual obsolescencia. Finalmente, se señalan riesgos reputacionales vinculados con la pérdida de confianza pública y la generación de controversias. Para cada uno de estos escenarios, se han delineado mitigaciones preliminares que orientan la gestión proactiva de los riesgos identificados.

Sección 8: Métricas de Éxito e Indicadores de Impacto - La medición del desempeño del sistema se estructura en torno a indicadores clave (KPIs) que abarcan aspectos técnicos, de servicio y de cumplimiento normativo. En el ámbito técnico, se consideran métricas como la precisión de los resultados, la disponibilidad operativa y los tiempos de respuesta. En relación con el impacto en el servicio, se incluyen indicadores de eficiencia institucional, niveles de satisfacción de los usuarios y cobertura poblacional alcanzada. Los KPIs de cumplimiento se centran en la ejecución de evaluaciones de riesgo algorítmico (ARA) y de impacto en protección de datos (DPIA), así como en la implementación de controles y auditorías periódicas. Para establecer una línea base comparativa, se documenta el estado actual de los procesos, lo que permite medir mejoras de manera objetiva. Asimismo, se definen metas cuantificables a seis y doce meses, con el propósito de orientar la toma de decisiones y evaluar el progreso del sistema en función de sus objetivos estratégicos.

Sección 9: Plan de Despliegue, Formación y Comunicación - La estrategia de implementación contempla diversas modalidades, entre las que se incluyen el desarrollo de pilotos, el despliegue gradual por fases y el enfoque de implementación total o *big bang*, según la naturaleza del sistema y el contexto institucional. El alcance inicial se define en función de las capacidades disponibles, mientras que el escalamiento se planifica conforme a criterios de madurez tecnológica y sostenibilidad operativa. En este marco, se han identificado necesidades específicas de capacitación tanto para funcionarios públicos como para ciudadanos, con el objetivo de asegurar una apropiada adopción y comprensión del sistema. Cuando el proyecto lo amerite, se establece un plan de comunicación y divulgación dirigido a la ciudadanía, orientado a promover la transparencia, la participación informada y la confianza pública. Finalmente, se presenta un cronograma estimado de implementación que permite visualizar las etapas clave del proceso y anticipar los recursos requeridos en cada fase.

Sección 10: Monitoreo, Auditoría y Respuesta a Incidentes - El sistema incorpora controles de monitoreo continuo que abarcan dimensiones técnicas, de equidad algorítmica y de privacidad, con el fin de garantizar su funcionamiento ético y conforme a los estándares normativos vigentes. La

frecuencia de las revisiones y auditorías se establece en función del nivel de riesgo asociado y de los requerimientos regulatorios aplicables. En caso de presentarse incidentes, se activa un protocolo de respuesta que define responsabilidades, tiempos de reacción y mecanismos de mitigación. Adicionalmente, cuando el sistema tiene implicaciones directas sobre la ciudadanía, se habilitan canales específicos para la recepción de reclamaciones, los cuales deben ser accesibles, eficaces y respetuosos de los derechos fundamentales.

Sección 11: Criterios y Plan de Fin de Vida - El sistema deberá ser retirado en caso de que se presenten condiciones como obsolescencia tecnológica, aparición de riesgos considerados inaceptables, una relación costo-beneficio desfavorable o modificaciones sustantivas en el marco regulatorio vigente. Ante tales escenarios, se establece un plan preliminar para la gestión de datos al final del ciclo de vida del sistema, el cual contempla medidas de resguardo, eliminación segura y cumplimiento de obligaciones legales en materia de protección de datos personales. Asimismo, se propone un plan de transición hacia soluciones alternativas, con el fin de garantizar la continuidad operativa y mitigar impactos negativos sobre los procesos institucionales afectados.

Sección 12: Aprobaciones y Decisión - La evaluación de viabilidad técnica queda a cargo del Responsable Técnico, quien debe determinar si el proyecto es viable, viable con condiciones o no viable. En paralelo, el área Jurídica realiza una evaluación de conformidad legal, clasificando la propuesta como conforme, sujeta a ajustes o no conforme. Por su parte, el Delegado de Protección de Datos (DPO) evalúa los aspectos de privacidad, indicando si el proyecto está aprobado, requiere la elaboración de una Evaluación de Impacto en la Protección de Datos (DPIA) o no ha sido aprobado. La dimensión presupuestal es revisada por el equipo de Planeación, que establece si existen recursos disponibles, si se requiere gestión adicional o si el proyecto resulta no viable desde el punto de vista financiero. Finalmente, el Comité de IA emite una decisión formal, la cual puede consistir en la aprobación para proceder a la Fase 2, la aprobación con condiciones, el rechazo o la solicitud de información adicional. Esta decisión debe ser refrendada mediante la firma del Presidente del Comité, acompañada de la fecha correspondiente.

4.2.2. Matriz de Riesgos de IA

El propósito de esta herramienta es ofrecer un instrumento estandarizado que facilite la identificación, evaluación y priorización de riesgos asociados al uso de sistemas basados en IA. Esta herramienta se fundamenta en un enfoque analítico que considera tanto la probabilidad de ocurrencia como el impacto potencial de cada riesgo, evaluados a través de múltiples dimensiones críticas para el entorno institucional.

La metodología adoptada se estructura en una matriz de evaluación de 3x3, en la cual se cruzan los niveles de probabilidad e impacto, clasificados en tres categorías: bajo, medio y alto. Esta combinación genera una calificación de riesgo final que oscila entre 1 y 9, permitiendo su categorización en tres niveles: bajo (1-3), medio (4-6) y alto (7-9). Cada nivel de riesgo conlleva un conjunto específico de acciones requeridas, orientadas a mitigar sus posibles efectos adversos. En este sentido, la matriz no solo actúa como un mecanismo de diagnóstico, sino también como una guía para la toma de decisiones estratégicas en la gestión responsable de tecnologías basadas en IA.

El manual de usuario correspondiente a esta herramienta se encuentra en el Anexo F2.

Dimensiones de impacto

La evaluación de riesgos en sistemas de inteligencia artificial debe considerar múltiples dimensiones que reflejan el alcance potencial de sus consecuencias. En primer lugar, se encuentra la dimensión relativa a los derechos y libertades fundamentales, la cual contempla el riesgo de que los sistemas vulneren o restrinjan garantías constitucionales como la privacidad, la igualdad ante la ley, el debido proceso, la libertad de expresión y el acceso equitativo a servicios públicos. Esta dimensión resulta crítica, dado que cualquier afectación en estos ámbitos puede comprometer la legitimidad institucional y la protección de los derechos humanos (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021).

En segundo lugar, la equidad y la no discriminación constituyen una dimensión clave para identificar sesgos algorítmicos, así como posibles mecanismos de exclusión que afecten a grupos protegidos o en situación de vulnerabilidad.

La tercera dimensión corresponde a la continuidad y calidad del servicio, entendida como el impacto que los sistemas de IA pueden tener sobre la capacidad institucional para garantizar la prestación de servicios esenciales. Aspectos como la disponibilidad, la confiabilidad operativa y la resiliencia frente a fallos técnicos son elementos centrales en esta evaluación (Organisation for Economic Co-operation and Development, 2022).

La cuarta dimensión aborda la seguridad y la ciberseguridad, considerando la exposición a ataques maliciosos, brechas de datos sensibles y fallos que puedan derivar en consecuencias físicas o digitales de alto riesgo. En este contexto, la protección de infraestructuras críticas y la integridad de los sistemas se convierten en prioridades estratégicas (European Commission, 2021).

En quinto lugar, se analiza el impacto sobre la reputación institucional y la confianza pública. El uso de tecnologías opacas o mal gestionadas puede deteriorar la percepción ciudadana respecto a la

transparencia y responsabilidad del sector público, generando resistencia social y pérdida de legitimidad (Floridi et al., 2018).

Finalmente, la dimensión de cumplimiento regulatorio contempla el riesgo de que los sistemas de IA infrinjan normativas vigentes, lo cual podría acarrear sanciones legales, responsabilidades administrativas o consecuencias reputacionales. Esta dimensión exige una revisión constante del marco jurídico aplicable y una articulación efectiva entre los equipos técnicos y jurídicos (González Fuster, 2020)

Escala de probabilidad e impacto

La evaluación de riesgos en sistemas de inteligencia artificial requiere una estimación sistemática tanto de la probabilidad de ocurrencia como del impacto potencial de cada evento identificado. Para ello, la matriz propone una escala ordinal de tres niveles —bajo, medio y alto— que permite clasificar de forma estandarizada los distintos escenarios de riesgo.

En cuanto a la probabilidad, se establecen tres categorías. El nivel bajo (valor 1) se asigna a eventos cuya ocurrencia se considera improbable, es decir, con una probabilidad inferior al 10 % durante el horizonte temporal previsto para la operación del sistema. El nivel medio (valor 2) corresponde a situaciones con una posibilidad razonable de materialización, estimada entre el 10 % y el 50 %. Finalmente, el nivel alto (valor 3) se reserva para aquellos riesgos cuya ocurrencia se considera probable o altamente probable, superando el umbral del 50 %.

Por su parte, la escala de impacto también se estructura en tres niveles. El impacto bajo (valor 1) se refiere a consecuencias menores, de alcance localizado y reversibles con un esfuerzo mínimo. El impacto medio (valor 2) implica efectos significativos que, aunque no necesariamente irreversibles, requieren un esfuerzo sustancial para su mitigación y pueden afectar a grupos específicos de personas. En contraste, el impacto alto (valor 3) se asocia con consecuencias graves, difíciles o imposibles de revertir, que comprometen derechos fundamentales o afectan a un número considerable de individuos, además de generar un daño reputacional severo para la entidad responsable.

La combinación de estos dos ejes —probabilidad e impacto— permite calcular una calificación de riesgo mediante la multiplicación de ambos valores, generando un rango que va de 1 a 9. Esta puntuación se traduce en tres niveles de riesgo: bajo (1 a 3), medio (4 a 6) y alto (7 a 9), los cuales orientan la selección de medidas de mitigación proporcionales a la magnitud del riesgo identificado.

Plantilla de la matriz y acciones según nivel de riesgo

La Matriz de Riesgos de IA se operacionaliza mediante una plantilla estructurada que permite documentar de forma sistemática cada riesgo identificado. Esta plantilla incluye los siguientes campos:

ID del riesgo, descripción detallada, categoría temática, dimensiones de impacto involucradas, probabilidad estimada (valor entre 1 y 3), impacto proyectado (valor entre 1 y 3), calificación total obtenida por la multiplicación de probabilidad e impacto, nivel de riesgo resultante, controles existentes, evaluación de la efectividad de dichos controles, controles adicionales propuestos, responsable asignado y estado actual del riesgo.

Este formato facilita la trazabilidad de los riesgos y la articulación de medidas de mitigación proporcionales a su nivel. La calificación final, obtenida mediante la fórmula P × I, permite clasificar el riesgo en tres niveles: bajo (1 a 3), medio (4 a 6) y alto (7 a 9). Cada nivel conlleva un conjunto diferenciado de acciones institucionales.

Para los riesgos clasificados como bajos, se recomienda mantener un monitoreo rutinario, aplicar controles básicos y realizar una revisión anual del estado del riesgo. En el caso de los riesgos medios, se requiere la implementación de controles específicos, un monitoreo más frecuente —mensual o trimestral—, revisión semestral y aprobación por parte de niveles gerenciales. Finalmente, los riesgos altos demandan una respuesta más robusta: deben ser mitigados antes del despliegue del sistema, contar con controles reforzados obligatorios, someterse a monitoreo continuo y revisión frecuente. Además, pueden requerir el rediseño del sistema o incluso el rechazo del caso de uso, y su aprobación debe ser otorgada por la Alta Dirección o el Comité de IA correspondiente.

Este enfoque escalonado permite una gestión proporcional del riesgo, alineada con principios de gobernanza tecnológica responsable, y asegura que los sistemas de IA operen dentro de márgenes aceptables de seguridad, equidad y legalidad.

4.2.3. Plantilla ARA / DPIA

La plantilla ARA/DPIA constituye uno de los instrumentos más críticos dentro del toolkit del framework, al integrar en un solo documento la *Evaluación de Riesgos Algorítmicos (ARA)* y la *Evaluación de Impacto en Protección de Datos (DPIA)*. Su propósito es ofrecer una metodología estructurada que permita identificar, analizar y mitigar los riesgos derivados del uso de sistemas de inteligencia artificial, tanto en lo referente a derechos fundamentales como a aspectos de privacidad, equidad, seguridad y transparencia.

Este instrumento es fundamental para asegurar la alineación con marcos regulatorios como la **Circular 002 de la SIC**, el **CONPES 4144 (2024)**, la **Ley 1581 de Habeas Data (2012)**, el **AI Act**, así como con estándares internacionales como **ISO/IEC 23894**, **ISO 42001**, **NIST AI RMF** y buenas prácticas de gobernanza algorítmica. La Figura 14 muestra la estructura completa y las doce secciones que componen la plantilla ARA/DPIA, herramienta central para la evaluación integral de riesgos e impactos de los sistemas de IA.

El manual de usuario correspondiente a esta herramienta se encuentra en el Anexo F3.

Sección 1. Información General

Esta sección reúne los datos básicos que identifican el sistema evaluado, tales como la entidad responsable, la unidad operativa, la denominación del sistema, los responsables del análisis, la fecha de elaboración y el número de versión.

Su función principal es establecer trazabilidad documental y permitir que cualquier auditor o evaluador comprenda quién está a cargo, cuándo se elaboró el documento y sobre qué sistema recae la evaluación.
Se completa al inicio y se actualiza cada vez que se modifique el modelo o el análisis de riesgos. Esta sección establece los datos fundamentales de identificación y trazabilidad del ARA/DPIA.

Sección 2. Descripción del Sistema y Alcance

Incluye una explicación clara del sistema de IA: su propósito, los elementos técnicos que lo componen, la base legal que justifica su implementación, la población a la que afecta, los casos de uso permitidos y los expresamente prohibidos, así como el tipo de decisiones que genera.
Esta sección permite comprender de manera integral qué hace el sistema y en qué condiciones debe operar. También ayuda a prevenir usos indebidos, sobre todo cuando un sistema podría ampliarse a contextos no autorizados.

Se completa con apoyo del responsable técnico y del área jurídica. La sección está diseñada para capturar de manera exhaustiva las características fundamentales y los límites operativos del sistema de IA bajo evaluación.

Sección 3. Datos y Origen de la Información

La tercera sección integra el mapeo de datos y una versión simplificada del Data Sheet. En ella se describen las categorías de información utilizadas, el tipo de dato, las fuentes de origen, la licitud del tratamiento, las técnicas aplicadas, los sesgos potenciales y las observaciones relevantes. Con esta información se evalúan aspectos fundamentales como calidad, representatividad, presencia de datos personales y eventuales riesgos de uso indebido.
La sección es crítica para dar cumplimiento a la normativa de protección de datos y para anticipar posibles sesgos o desigualdades.

Se completa en conjunto con el equipo de datos. Esta sección documenta de manera sistemática el mapeo de datos, su procedencia y las características relevantes para evaluar riesgos de calidad y privacidad.

Sección 4. Base Legal y Consentimiento

Aquí se identifica la base jurídica que permite el tratamiento de los datos, incluyendo si el sistema requiere o no consentimiento explícito de los titulares. También se documentan los mecanismos de obtención del consentimiento y las alternativas aplicables cuando este no es viable.
La finalidad de esta sección es demostrar que el tratamiento de datos es legítimo y compatible con la legislación vigente.

Generalmente la llena el área jurídica y el Delegado de Protección de Datos. En esta sección se registra de forma clara los soportes jurídicos del tratamiento de datos y los mecanismos de consentimiento aplicables.

Sección 5. Evaluación de Impactos en Derechos

Esta sección analiza los posibles impactos del sistema sobre cuatro derechos fundamentales: privacidad, igualdad, debido proceso y acceso a servicios públicos.

Para cada derecho se responde a una pregunta orientadora, se describe el riesgo asociado, y se valora la probabilidad, el impacto y el nivel de riesgo (bajo, medio o alto).

Es una de las secciones más importantes del DPIA, ya que determina si el sistema puede afectar derechos fundamentales y si se requieren medidas adicionales de mitigación.

En esta sección se analiza de manera sistemática el efecto potencial del sistema sobre derechos fundamentales como la privacidad, la igualdad, el debido proceso y el acceso a servicios.

Sección 6. Matriz de Riesgos Algorítmicos

La matriz consolida los riesgos identificados y asigna una puntuación basada en la probabilidad y el impacto. Cada riesgo se clasifica por categoría (por ejemplo, privacidad, equidad o seguridad) y se registran los controles existentes.

Esta sección facilita la priorización de riesgos y permite tomar decisiones informadas sobre la viabilidad del sistema.

Esta sección proporciona una herramienta estructurada para consolidar, evaluar y priorizar los riesgos identificados durante el análisis.

Sección 7. Medidas de Mitigación y Controles

Una vez definidos los riesgos, en esta sección se especifican las medidas para atenderlos: de qué tipo son (técnicas u organizativas), quién es el responsable, en qué plazo se implementarán, cuál es su indicador de cumplimiento y cuál es su estado actual.

Es, en la práctica, el plan de acción que la entidad debe seguir para reducir los riesgos identificados. Se revisa y actualiza con periodicidad. En esta sección se registra el plan de acción específico para cada riesgo, incluyendo responsables, plazos e indicadores de seguimiento.

Sección 8. Supervisión Humana

Esta sección define los mecanismos mediante los cuales los seres humanos supervisarán y podrán intervenir en el funcionamiento del sistema. Incluye los criterios de escalamiento, los roles responsables, la frecuencia de revisión y los pasos a seguir en caso de desacuerdo entre la decisión humana y la decisión del sistema.

Su propósito es asegurar que toda decisión automatizada esté respaldada por supervisión humana efectiva.

Corresponde al cumplimiento del principio de "intervención humana significativa". Esta sección documenta los protocolos y responsabilidades para garantizar una supervisión humana efectiva y significativa sobre las decisiones automatizadas.

Sección 9. Monitoreo Continuo y KPI

El monitoreo continuo recoge los indicadores con los que se evaluará periódicamente el desempeño del sistema. Se definen las métricas, su fórmula, el valor objetivo, el valor actual, la frecuencia de medición y el responsable de cada una.

Permite identificar degradación del modelo (drift), cambios en los datos, variaciones de desempeño y efectos no deseados. Esta sección estructura el plan de seguimiento mediante indicadores clave para evaluar el desempeño y la estabilidad del sistema en operación.

Sección 10. Comunicación y Transparencia

Aquí se documentan los mecanismos para informar a los ciudadanos y los usuarios del sistema acerca del uso de la inteligencia artificial, así como los documentos publicados y los canales para consultas o reportes de problemas.

La sección permite demostrar que la entidad cumple los principios de transparencia y rendición de cuentas.

Es especialmente importante cuando el sistema afecta a ciudadanos de manera directa. Esta sección registra las estrategias y canales establecidos para informar a los ciudadanos y garantizar la transparencia en el uso de sistemas de IA.

Sección 11. Auditoría y Actualizaciones

Registra las auditorías internas o externas realizadas al sistema, su alcance, frecuencia, fecha de próxima evaluación, responsables, hallazgos y acciones correctivas.
La finalidad es asegurar el seguimiento permanente del sistema y el cumplimiento de las medidas de calidad y seguridad. Esta sección documenta el programa de auditorías, sus hallazgos y las acciones correctivas derivadas para garantizar la mejora continua del sistema.

Sección 12. Aprobaciones y Decisión Final

Es la sección que formaliza la decisión institucional. Incluye la evaluación del Delegado de Protección de Datos, del responsable técnico y del área jurídica.

Finalmente, se consigna la decisión del Comité de IA, las condiciones establecidas, la fecha de aprobación y la fecha programada para la siguiente revisión.

Sin esta sección firmada, el sistema no debería entrar en operación. Esta sección consolida las validaciones institucionales requeridas y formaliza la decisión de implementación del sistema.

Importancia dentro del Framework

Esta herramienta resulta esencial en el marco de gobernanza porque permite cumplir con los requerimientos regulatorios colombianos, como la Circular 002 y la Ley 1581 (2012), así como con estándares internacionales, entre ellos el AI Act. Su implementación garantiza que los sistemas de inteligencia artificial no vulneren derechos fundamentales, además de proveer un marco unificado que evita la creación de formatos dispares por parte de cada entidad. Asimismo, facilita la trazabilidad y la auditoría del ciclo de vida algorítmico, aportando criterios objetivos para aprobar o rechazar sistemas de IA en el sector público. De esta manera, contribuye a reducir riesgos reputacionales y operativos mediante la anticipación de posibles fallos.

Este instrumento constituye la pieza central del enfoque de gobernanza, ya que articula de manera secuencial los componentes clave: datos, modelos, riesgos, medidas de mitigación, supervisión, transparencia, auditoría y aprobación final.

¿Cómo se utiliza la plantilla?

La plantilla está diseñada para aplicarse de forma secuencial y colaborativa por diferentes áreas: técnica, jurídica, delegados de protección de datos, comité de IA, equipos de datos y analítica, así como equipos de TI y seguridad. El proceso de uso comprende las siguientes etapas:

1. El área técnica completa la descripción del sistema y los datos utilizados.
2. El equipo jurídico valida la base legal y los mecanismos de consentimiento.

3. El delegado de protección de datos analiza riesgos de privacidad y determina la necesidad de una evaluación completa (DPIA).
4. El equipo multidisciplinario evalúa riesgos algorítmicos y llena la matriz correspondiente.
5. Se documentan los controles y medidas de mitigación.
6. Los equipos de TI y seguridad definen indicadores clave (KPIs) y procedimientos de monitoreo.
7. Se establecen protocolos de transparencia y mecanismos de participación ciudadana.
8. Auditoría revisa la consistencia del instrumento.
9. Finalmente, el comité de IA toma la decisión de aprobación.

¿Por qué es clave para el sector público?

La ARA/DPIA proporciona transparencia algorítmica, responsabilidad institucional y evidencia de cumplimiento normativo. Además, ofrece protección jurídica tanto para la entidad como para los funcionarios, asegura trazabilidad para auditorías internas y externas, y establece un estándar unificado aplicable a todas las entidades del Distrito. En síntesis, esta herramienta garantiza que ningún sistema de IA se despliegue sin un análisis riguroso de riesgos y derechos.Conclusión

Esta plantilla se convierte en el instrumento rector de gobernanza de IA, articulando datos, modelos, riesgos, legalidad, derechos fundamentales y transparencia. Su diseño facilita la adopción del framework por entidades con capacidades técnicas diversas y garantiza alineación con los estándares nacionales e internacionales.

La plantilla se convierte en el instrumento rector de la gobernanza de IA, articulando datos, modelos, riesgos, legalidad, derechos fundamentales y transparencia. Su diseño facilita la adopción del framework por entidades con capacidades técnicas diversas y asegura la alineación con los estándares nacionales e internacionales.

4.2.4. Data Sheet

El *Data Sheet* constituye un instrumento de documentación diseñado para describir de manera exhaustiva las características, la calidad, el origen, la composición, los riesgos y las limitaciones del conjunto de datos utilizado para entrenar, validar o ejecutar un sistema de inteligencia artificial. Su propósito principal es garantizar que las entidades públicas comprendan el funcionamiento, la procedencia y las implicaciones del dataset antes de su incorporación en cualquier proceso automatizado. Esta herramienta, propuesta inicialmente por investigadores del MIT como un estándar para la transparencia de datos (Gebru et al., 2018), se ha consolidado como un componente esencial dentro de los marcos de gobernanza y en regulaciones internacionales, tales como el *NIST AI RMF*, las

normas ISO/IEC 25012 e ISO/IEC 5259, así como en diversas políticas públicas de IA. En el contexto del Distrito, el *Data Sheet* asegura que los datos utilizados sean confiables, legales, representativos y compatibles con criterios éticos y de protección de derechos fundamentales.

Alineación con la Circular 002 de la SIC

La Circular 002 establece que cualquier entidad que procese datos debe documentar su origen, registrar su finalidad, evaluar riesgos sobre los titulares, identificar datos personales y sensibles, y definir medidas de seguridad y minimización (Superintendencia de Industria y Comercio [SIC], 2023). El *Data Sheet* se ajusta plenamente a estos requerimientos, ya que documenta el flujo de datos y su procedencia, justifica la base legal del tratamiento, identifica datos sensibles y su necesidad, evalúa riesgos para los titulares, registra controles de seguridad y permite auditorías sobre el uso del dataset. Por ello, se convierte en una herramienta indispensable para demostrar el cumplimiento normativo ante la SIC.

Alineación con el CONPES 4144

El CONPES 4144 establece que los sistemas de IA en el país deben garantizar gobernanza de datos, trazabilidad del ciclo de vida, identificación de sesgos, transparencia en los procesos, uso responsable, protección de derechos fundamentales y mitigación de riesgos (Departamento Nacional de Planeación, 2022). El *Data Sheet* responde directamente a estos lineamientos porque documenta la calidad, representatividad y sesgos del conjunto de datos, permite identificar discriminación estructural, facilita la trazabilidad completa del dataset, registra transformaciones, limpieza y medidas de seguridad, define usos permitidos e indebidos, y evalúa riesgos éticos y sociales. En consecuencia, se configura como la “hoja de vida” del dataset y como un mecanismo de control esencial dentro del framework de gobernanza.

El manual de usuario correspondiente a esta herramienta se encuentra en el Anexo F4.

Descripción y estructura de la herramienta

La herramienta se organiza en trece secciones que documentan los componentes esenciales del dataset:

Sección 1: Información general del dataset

Incluye nombre, entidad propietaria, área responsable, versión, fecha de creación/actualización y contacto institucional. Permite trazabilidad administrativa. Esta sección recopila los datos básicos de identificación y gestión administrativa del conjunto de datos, fundamentales para su trazabilidad y referencia institucional.

Sección 2: Descripción y propósito

Explica qué contiene el dataset y cuál es su finalidad dentro del proceso institucional o servicio público que soporta. Esta sección define de manera clara el contenido del conjunto de datos y su función específica dentro de los procesos o servicios públicos que utiliza.

Sección 3: Origen y método de recolección

Describe las fuentes (sistemas internos, sensores, encuestas, terceros autorizados), la metodología de captura y la frecuencia de actualización.

Sección 4: Composición del dataset

Incluye número de registros, número de variables, tipos de datos, descripción del diccionario de datos y representatividad poblacional. Aquí se documentan las características estructurales y estadísticas fundamentales del conjunto de datos, proporcionando una visión clara de su escala y diversidad.

Sección 5: Presencia de datos personales y sensibles

Clasifica los datos en personales, sensibles o no personales, justifica su uso y especifica la base legal del tratamiento. En esta sección se evalúa y documenta la naturaleza de la información contenida en el dataset, asegurando el cumplimiento de la normativa de protección de datos personales y sensibles.

Sección 6: Calidad del dataset

Evalúa datos faltantes, inconsistencias, duplicados, procesos de limpieza, técnicas de validación y estándares de calidad aplicados. En esta sección sistematiza la evaluación de la integridad, consistencia y confiabilidad de los datos mediante métricas y procesos documentados de limpieza y validación.

Sección 7: Evaluación de sesgos y representatividad

Analiza sesgos potenciales, distribución de variables sensibles y riesgos para poblaciones vulnerables. Esta sección documenta los análisis realizados para identificar posibles desbalances, subrepresentaciones o sesgos sistemáticos en el conjunto de datos que puedan afectar la equidad de los modelos.

Sección 8: Procesamiento y transformaciones aplicadas

Documenta normalización, imputación, balanceo, anonimización o cualquier operación realizada sobre los datos. Aquí se registra de manera detallada las operaciones de preparación, limpieza y transformación ejecutadas sobre los datos antes de su uso en modelos de IA.

Sección 9: Riesgos éticos y legales

Identifica riesgos de discriminación, reidentificación, uso indebido, sesgos estructurales y vulneración de derechos. En esta sección se documenta la identificación de amenazas potenciales relacionadas con el uso del conjunto de datos, desde perspectivas tanto éticas como de cumplimiento normativo.

Sección 10: Uso permitido y no permitido

Establece condiciones de uso, finalidades autorizadas, limitaciones y restricciones de acceso. Aquí se define claramente los límites operativos y éticos del conjunto de datos, especificando los contextos de uso aprobados y aquellos expresamente prohibidos.

Sección 11: Seguridad del dataset

Incluye medidas de cifrado, accesos permitidos, políticas de custodia y controles de seguridad. En esta sección se documenta las salvaguardas técnicas y organizativas implementadas para proteger la integridad, confidencialidad y disponibilidad del conjunto de datos.

Sección 12: Historial del dataset

Registra versiones, cambios realizados, fecha de modificación y responsable del ajuste. En esta sección se mantiene un registro cronológico de todas las modificaciones y actualizaciones realizadas al conjunto de datos, garantizando trazabilidad completa de su evolución.

Sección 13: Aprobaciones institucionales

Incluye validación del equipo de datos, área jurídica y Comité Distrital de IA. Aquí se consolidan las validaciones requeridas de los diferentes actores responsables, formalizando la aprobación del conjunto de datos para su uso en sistemas de IA.

Importancia dentro del framework

El *Data Sheet* desempeña un papel esencial en el marco de gobernanza de inteligencia artificial, ya que contribuye a eliminar la opacidad en la recolección y manipulación de datos, previene que los modelos se entrenen sobre información discriminatoria y permite la realización de auditorías tanto internas como externas. Además, garantiza el cumplimiento legal, protege a los ciudadanos y a las poblaciones vulnerables, facilita la rendición de cuentas y estandariza los procesos relacionados con la calidad y la gobernanza de datos. Junto con el Diccionario de Datos y la *Model Card*, conforma un sistema integral orientado a la transparencia algorítmica, lo que refuerza la confianza institucional y asegura la trazabilidad en todo el ciclo de vida del modelo.

4.2.5. Checklist de Evaluación de Proveedores de IA

El presente Checklist constituye una herramienta operativa fundamental del Framework de Gobernanza de IA, diseñada para estandarizar y robustecer los procesos de selección y contratación de proveedores de sistemas de inteligencia artificial en el Distrito Capital. Su función principal es servir como instrumento de debida diligencia, permitiendo a las entidades públicas evaluar de manera objetiva, sistemática y proporcional al riesgo, la capacidad de un proveedor para cumplir con los requisitos normativos, técnicos y éticos establecidos. El checklist evalúa integralmente siete dimensiones críticas: 1) la conformidad con el marco regulatorio colombiano como la Ley 1581 (2012) y CONPES 4144 (2024) y la gobernanza interna del proveedor; 2) la transparencia y documentación técnica de los modelos y datos; 3) las garantías de privacidad y protección de datos personales; 4) la seguridad, robustez y resiliencia de los sistemas; 5) las capacidades de auditoría y rendición de cuentas; 6) la calidad del servicio y el soporte técnico; y 7) la sostenibilidad de la solución. Al aplicar este instrumento, las entidades no solo mitigan riesgos legales, reputacionales y operativos, sino que también fomentan un mercado de proveedores más maduro y alineado con los principios de una IA responsable en el sector público. Se proponen **tres versiones** adaptadas al nivel de riesgo.

El manual de usuario correspondiente a esta herramienta se encuentra en el Anexo F5.

En lugar de un checklist único y masivo, proponemos **tres versiones**:

Checklist Básico (Para Riesgo Mínimo/Limitado): 15 criterios críticos. Enfocado en privacidad según la Ley 1581 (2012), seguridad básica y funcionalidad. Se aplica en procesos de menor cuantía o para herramientas de productividad interna.

Checklist Estándar (Para Riesgo Alto): El checklist actual, pero rationalizado a ~25 criterios. Es la versión completa que se incluiría en pliegos tipo para la mayoría de licitaciones públicas.

Checklist Reforzado (Para Riesgo Inaceptable o Crítico): El checklist estándar + criterios adicionales de auditoría profunda, pruebas de estrés ético específicas y cláusulas contractuales excepcionales (ej.: para sistemas de biometría o salud).

Esto reduce la carga administrativa donde el riesgo lo permite.

Instrucciones Generales:

Seleccione el Nivel de Checklist según la clasificación de riesgo del caso de uso:

BÁSICO: Para sistemas de **Riesgo Mínimo o Limitado** (ej.: chatbots informativos, herramientas de productividad interna).

ESTÁNDAR: Para sistemas de **Alto Riesgo** (ej.: priorización de trámites, evaluación de beneficiarios, sistemas de recomendación que afecten acceso a servicios). *Este es el checklist por defecto para la mayoría de licitaciones.*

REFORZADO: Para sistemas de **Riesgo Inaceptable o Crítico** (ej.: biometría, sistemas de salud, puntuación social). Incluye todos los criterios del Estándar más requisitos excepcionales.

Para cada criterio, marque la opción que mejor describa la evidencia presentada por el proveedor. Utilice la columna "Evidencia/Referencia" para anotar el documento o sección específica que respalda su evaluación.

Sección 1: Conformidad Regulatoria Y Gobernanza (Peso: 25%)

Esta sección evalúa el grado en que el proveedor cumple con las regulaciones colombianas de protección de datos establecidas en la Ley 1581 (2012), así como la existencia de políticas formales orientadas a una inteligencia artificial responsable. También considera la implementación de sistemas de gestión de IA, como la norma ISO/IEC 42001, y la adopción de procesos sistemáticos para identificar y gestionar riesgos asociados al uso de inteligencia artificial.

Tabla 4. Checklist Evaluación Proveedores: Sección 1

ID	Criterio	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia
1.1	**Conformidad con marco regulatorio colombiano**	No tiene política de privacidad	Política básica no alineada	Política robusta y alineada con Ley 1581 (2012)		
1.2	**Política de IA Responsable o Ética**	No tiene política documentada	Borrador o política interna no formal	Política pública formal y referenciable		
1.3	**Sistema de Gestión de IA (AIMS)**	No tiene sistema de gestión	Elementos de un sistema pero no formalizado	Sistema documentado e implementado (ej. ISO/IEC 42001)		
1.4	**Gestión de Riesgos de IA**	No tiene proceso definido	Evaluaciones de riesgo ad-hoc	Proceso sistemático y documentado		

Fuente: Elaboración Propia

Sección 2: Documentación Y Transparencia Técnica (Peso: 20%)

Se verifica que el proveedor proporcione documentación técnica completa, incluyendo Model Cards que describen métricas y limitaciones del modelo y Data Sheets con información detallada sobre los

datos de entrenamiento. Además, se exige documentación auditable que permita comprender la arquitectura del sistema, los hiperparámetros utilizados y las decisiones clave en el diseño.

Tabla 5. Checklist Evaluación Proveedores: Sección 2

ID	Criterio	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia
2.1	**Model Card (Ficha del Modelo)**	No proporciona Model Card	Model Card básico, falta info clave	Model Card completo		
2.2	**Data Sheet (Ficha de Datos)**	No proporciona Data Sheet	Descripción básica de datos	Data Sheet detallado		
2.3	**Documentación Técnica para Auditoría**	No proporciona documentación	Documentación superficial	Documentación detallada y bitácoras		

Fuente: Elaboración Propia

Sección 3: Privacidad Y Protección De Datos (Peso: 20%)

En esta sección se examina la claridad contractual respecto a la titularidad de los datos, la existencia de acuerdos sólidos de procesamiento (DPA) alineados con la Ley 1581 (2012) y los procedimientos que garantizan el ejercicio de derechos de habeas data, como acceso, rectificación y supresión. También se revisan las políticas específicas de retención y eliminación de información.

Tabla 6. Checklist Evaluación Proveedores: Sección 3

ID	Criterio	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia
3.1	**Claridad sobre Titularidad de Datos**	No hay claridad contractual	Titularidad definida con ambigüedades	Contrato define claramente la titularidad		
3.2	**Data Processing Agreement (DPA)**	No ofrece un DPA	DPA básico que requiere ajustes	DPA robusto y alineado con Ley 1581 (2012)		
3.3	**Gestión de Derechos de los Titulares**	No tiene procedimientos	Procedimientos manuales o lentos	Procedimientos eficientes y canales claros		
3.4	**Políticas de Retención y Borrado de Datos**	No tiene políticas definidas	Políticas genéricas	Políticas específicas y alineadas		

Fuente: Elaboración Propia

Sección 4: Seguridad Y Robustez (Peso: 20%)

Se evalúa la vigencia de certificaciones de seguridad como ISO 27001 y SOC 2, la realización de pruebas de robustez frente a ataques adversarios y datos fuera de distribución, así como la existencia de protocolos actualizados y comprobados para responder ante incidentes de seguridad.

Tabla 7. Checklist Evaluación Proveedores: Sección 4

ID	Criterio	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia
4.1	**Certificaciones de Seguridad**	No tiene certificaciones	Certificaciones básicas o con hallazgos	ISO 27001 o equivalente vigente		
4.2	**Pruebas de Robustez y Seguridad de IA**	No ha realizado pruebas de robustez	Pruebas básicas de rendimiento	Pruebas de robustez y seguridad con resultados		
4.3	**Respuesta a Incidentes de Seguridad**	No tiene protocolo documentado	Protocolo básico no probado	Protocolo completo, actualizado y probado		

Fuente: Elaboración Propia

Sección 5: Auditoría Y Rendición De Cuentas (Peso: 10%)

Esta sección verifica que el proveedor acepte cláusulas contractuales que otorguen derecho a auditoría con acceso completo a documentación y evidencias. Asimismo, se exige la implementación de sistemas de trazabilidad que incluyan registros comprensivos e inmutables de todas las decisiones tomadas por el sistema de IA.

Tabla 8. Checklist Evaluación Proveedores: Sección 5

ID	Criterio	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia
5.1	**Derecho a Auditoría**	Se niega a incluir cláusulas de auditoría	Acepta auditorías con limitaciones	Acepta cláusulas de derecho a auditoría		
5.2	**Trazabilidad de Decisiones**	No registra logs o son insuficientes	Registra logs básicos	Registra logs exhaustivos e inmutables		

Fuente: Elaboración Propia

Sección 6: Calidad Del Servicio Y Soporte (Peso: 15%)

Finalmente, se revisa la existencia de acuerdos de nivel de servicio (SLA) robustos, con métricas cuantificables y penalizaciones claras, la disponibilidad de soporte técnico prioritario y planes de transferencia de conocimiento. También se considera la transparencia en el roadmap, incluyendo la notificación anticipada de cambios significativos.

Tabla 9. Checklist Evaluación Proveedores: Sección 6

ID	Criterio	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia
6.1	**Acuerdos de Nivel de Servicio (SLA)**	No ofrece SLA cuantificables	SLA con métricas básicas sin penalizaciones	SLA robustos con métricas y compensaciones		
6.2	**Soporte Técnico y Transferencia de Conocimiento**	No ofrece plan de soporte	Soporte estándar y documentación básica	Soporte prioritario y plan de transferencia		
6.3	**Gestión de Cambios y Roadmap**	No comparte roadmap	Roadmap de alto nivel	Roadmap detallado y notificaciones >30 días		

Fuente: Elaboración Propia

Informe Final De Evaluación

El informe final de evaluación resume la revisión integral del proveedor en relación con criterios regulatorios, técnicos y contractuales, verificando el cumplimiento obligatorio de la Ley 1581 (2012), la capacidad de suscribir acuerdos de procesamiento de datos (DPA), la existencia de certificaciones de seguridad y la oferta de SLA con métricas claras. Con base en la puntuación ponderada y la verificación de estos requisitos, se determina si el proveedor es recomendado, recomendado con condiciones o no recomendado, incorporando observaciones finales y la validación por parte del responsable técnico, el área jurídica y el comité de IA. La Figura 55 presenta la estructura general del Checklist de Evaluación de Proveedores de IA, herramienta diseñada para estandarizar la evaluación de proveedores según su nivel de riesgo y capacidad de cumplimiento.

4.2.6. Model Card

La Model Card es un instrumento estándar internacional que permite documentar, de manera clara y estructurada, las características técnicas y operativas de los modelos de inteligencia artificial utilizados por una entidad pública. Originalmente propuesta por Google y actualmente adoptada por gobiernos, organismos multilaterales y estándares como el NIST AI RMF y la ISO/IEC 42001, su propósito es

garantizar que un modelo sea comprensible, auditable, explicable y supervisable a lo largo de su ciclo de vida.

Dentro del framework de Gobernanza de IA del Distrito, la Model Card cumple un rol central, ya que permite registrar información detallada sobre el propósito del modelo, su arquitectura, los datos utilizados, los riesgos identificados y los mecanismos de supervisión humana y monitoreo continuo. Al consolidar esta información, la herramienta facilita la trazabilidad del sistema, el reporte a los órganos de control y la rendición de cuentas ante la ciudadanía.

Alineación con la Circular 002 de la SIC

La Circular 002 establece obligaciones específicas para las entidades que implementan sistemas automatizados, entre las que se incluyen la documentación de dichos sistemas, la identificación de riesgos, la garantía de transparencia, la información clara al ciudadano sobre decisiones automatizadas y la demostración de supervisión humana. En este contexto, la *Model Card* constituye una herramienta que facilita el cumplimiento de estas disposiciones, ya que describe el modelo, explica su funcionamiento, documenta riesgos y sesgos, define los mecanismos de supervisión y permite informar de manera clara a los titulares de datos.

Alineación con el CONPES 4144

El documento CONPES 4144 (2024) establece lineamientos orientados a la transparencia, la responsabilidad, la gestión de riesgos, la equidad y la supervisión humana en el uso de tecnologías emergentes. La *Model Card* se alinea con estos principios porque ofrece trazabilidad del ciclo de vida del modelo, documenta métricas diferenciales para garantizar equidad, identifica riesgos éticos, técnicos y sociales, incorpora pautas para un uso responsable, describe mecanismos de supervisión y control, y facilita tanto auditorías internas como externas.

Descripción de la estructura de la herramienta

La herramienta está compuesta por catorce secciones diseñadas para capturar información esencial que asegure transparencia, seguridad y uso responsable del modelo. Cada sección cumple una función específica orientada a documentar aspectos técnicos, operativos y éticos, lo que permite una comprensión integral del sistema y su impacto.

El manual de usuario correspondiente a esta herramienta se encuentra en el Anexo F6.

Sección 1: Información general

Incluye nombre, versión, entidad, área responsable, fecha de creación, fecha de actualización y estado del ciclo de vida. Garantiza trazabilidad documental. Esta sección recopila los datos básicos de identificación y administración del modelo, esenciales para su trazabilidad y gestión documental.

Sección 2: Propósito del modelo

Describe el objetivo del modelo, su caso de uso, alcance, usuarios previstos y procesos institucionales impactados. Esta sección documenta de manera clara los objetivos, el alcance operativo y los impactos institucionales previstos para el sistema de IA.

Sección 3: Descripción técnica

Define el tipo de modelo (clasificación, regresión, NLP, etc.), la arquitectura, los algoritmos utilizados y las tecnologías empleadas. Esta sección documenta las características fundamentales de arquitectura, algoritmos y tecnologías que componen el modelo de IA.

Sección 4: Datos utilizados

Resume los datasets de entrenamiento, validación y prueba. Incluye referencias cruzadas al Data Sheet correspondiente. Esta sección documenta de manera organizada los conjuntos de datos empleados en el desarrollo y evaluación del modelo, asegurando trazabilidad con su documentación detallada.

Sección 5: Métricas de desempeño

Reporta métricas técnicas como precisión, recall, F1-score, AUC y métricas desagregadas por subpoblaciones. Permite evaluar equidad. Esta sección consolida los resultados cuantitativos del modelo, incluyendo tanto métricas globales como desgloses específicos para evaluar su rendimiento y equidad.

Sección 6: Evaluación de sesgos

Identifica sesgos potenciales, resultados de pruebas diferenciales y medidas aplicadas para mitigarlos. Esta sección documenta los análisis realizados para identificar sesgos potenciales y las medidas implementadas para promover la equidad del modelo.

Sección 7: Riesgos del modelo

Clasifica riesgos técnicos, operativos, sociales y éticos asociados al modelo y su implementación. Esta sección sistematiza la identificación y clasificación de los riesgos potenciales en diversas dimensiones, desde aspectos técnicos hasta impactos sociales y éticos.

Sección 8: Controles implementados

Documenta medidas técnicas y organizacionales, tales como validación humana, monitoreo, auditorías y alertas. En esta sección se registran las salvaguardas técnicas y organizativas establecidas para gestionar los riesgos identificados y asegurar la operación responsable del modelo.

Sección 9: Explicabilidad y transparencia

Describe técnicas de explicabilidad (SHAP, LIME, análisis de características, etc.), información pública y mecanismos de rendición de cuentas. En esta sección se documentan los métodos empleados para hacer comprensible el funcionamiento del modelo y los mecanismos establecidos para garantizar la transparencia ante usuarios y auditores.

Sección 10: Reglas de uso responsable

Define usos permitidos, usos restringidos, escenarios prohibidos y buenas prácticas operativas. En esta sección se establecen los límites operativos y éticos del modelo, especificando claramente los contextos de uso aprobados, restringidos y prohibidos.

Sección 11: Entradas y salidas del modelo

Especifica los tipos de datos aceptados por el modelo y el tipo de predicciones generadas. En esta sección se documentan las interfaces de datos del sistema, definiendo claramente qué información requiere para operar y qué tipo de resultados produce.

Sección 12: Monitoreo y mantenimiento

Indica la frecuencia de reentrenamiento, KPIs de calidad, métodos de detección de drift y responsables del monitoreo. En esta sección se establece el plan de vigilancia continua y las actividades de sostenimiento necesarias para garantizar el desempeño y la vigencia del modelo a lo largo de su ciclo de vida.

Sección 13: Historial de cambios

Registra modificaciones, fechas, responsables y justificaciones. En esta sección se documenta de manera cronológica todas las modificaciones realizadas al modelo, su documentación o sus parámetros, asegurando una trazabilidad completa de su evolución.

Sección 14: Aprobaciones

Incluye la validación del área técnica, jurídica, Delegado de Protección de Datos y Comité Distrital de IA. Esta sección consolida las firmas y validaciones institucionales requeridas para formalizar la implementación y el uso del modelo de IA, cerrando el ciclo de gobernanza documental.

Importancia dentro del framework

La herramienta adquiere un papel esencial en el marco de gobernanza de modelos de inteligencia artificial, ya que contribuye a evitar el uso de sistemas como "cajas negras", lo que incrementa la transparencia y la comprensión del proceso decisional. Además, facilita auditorías del sistema, reduce riesgos asociados a prácticas discriminatorias, documenta el cumplimiento normativo, y permite una supervisión humana efectiva en todas las etapas del ciclo de vida del modelo. De igual manera, protege a poblaciones vulnerables mediante la identificación temprana de sesgos, genera confianza ciudadana al garantizar procesos claros y verificables, y asegura la transparencia institucional frente a actores internos y externos. En síntesis, la *Model Card* se configura como la pieza clave para que el Distrito pueda operar modelos de IA de manera segura, ética y responsable, alineándose con los principios de gobernanza tecnológica contemporánea (CONPES, 2022; SIC, 2023).

4.2.7. Guía de Uso Interno de IA Generativa

La presente Guía Institucional de Uso Interno de IA Generativa establece los lineamientos, responsabilidades y protocolos necesarios para garantizar un uso seguro, ético y jurídicamente adecuado de las herramientas de inteligencia artificial generativa en las entidades públicas del Distrito Capital. Su propósito es ofrecer un marco claro y práctico que permita aprovechar los beneficios de estas tecnologías emergentes sin comprometer la protección de datos personales, la integridad institucional, la transparencia administrativa ni los derechos de la ciudadanía. Dado el acelerado avance de la GenAI y su creciente incorporación en el mapa de procesos misionales, estratégicos y de apoyo, esta guía se configura como un instrumento esencial para orientar su adopción responsable, alineado con la normativa colombiana vigente, las capacidades actuales del Distrito y las mejores prácticas internacionales adaptadas al contexto local.

4.2.7.1. Principales riesgos asociados al uso de IA Generativa en el sector público

Los sistemas de IA generativa presentan riesgos relevantes que deben gestionarse antes, durante y después de su uso institucional. Los principales riesgos identificados son:

Errores formales: La IA puede generar información incorrecta o inventada, lo cual puede afectar decisiones administrativas, comunicaciones oficiales o análisis para políticas públicas.

Riesgos de privacidad: Si el funcionario ingresa datos personales, reservados o clasificados, existe riesgo de filtración, tratamiento indebido o uso no autorizado. Esto puede constituir vulneración de la Ley 1581 (2012).

Sesgos y discriminación: Los modelos pueden reproducir sesgos históricos, afectando grupos vulnerables. Esto es especialmente crítico en entidades que atienden poblaciones específicas (niñez, mujeres, migrantes, personas mayores).

Dependencia tecnológica de proveedores privados: Puede generarse dependencia excesiva de plataformas de terceros, afectando la soberanía tecnológica y continuidad del servicio.

Riesgos reputacionales: La generación de textos erróneos, inadecuados o discriminatorios puede comprometer la imagen institucional.

Automatización sin supervisión: La ejecución automática de decisiones sin intervención humana puede generar fallas graves en trámites y servicios públicos.

4.2.7.2. Tipología de casos de uso relevantes en el Distrito Capital

En la Tabla 10 se presentan casos de uso iniciales y acotados, ajustados al contexto distrital:

Tabla 10. Guia de Uso Interno IA: Tipología de Casos de Uso Relevantes

Entidad / Dependencia	Caso de Uso Permitido	Nivel de Riesgo	Condiciones de Uso
Secretaría General	Redacción de borradores de documentos no sensibles	Bajo	Supervisión humana obligatoria
Sector Gobierno	Resúmenes preliminares de normativa o jurisprudencia	Bajo/Medio	Verificación jurídica posterior
DTI Distrital	Automatización de reportes técnicos o analítico	Medio	Validación de datos por expertos
Secretaría de Salud	Guías informativas al ciudadano	Bajo	No ingresar datos clínicos ni reservados
Secretaría de Movilidad	Chatbots informativos de normas y servicios	Medio/Alto	Pruebas piloto + Comité de IA
Uaesp / Ambiente	Análisis preliminar de datos abiertos	Bajo	Solo usar datos públicos
Jurídica Distrital	Borradores de conceptos no vinculantes	Bajo	Revisión posterior por abogado

Fuente: Elaboración Propia

4.2.7.3. Obligaciones mínimas del proveedor de IA generativa

Para plataformas internas o de terceros aplican los siguientes requisitos mínimos:

Protección de datos (Ley 1581 de 2012)

Prohibición contractual de almacenar prompts con datos personales.

Prohibición de entrenar modelos con información del Distrito.

Garantía de ubicación de datos en jurisdicciones compatibles (no exige nube nacional, pero sí cumplimiento estricto).

Seguridad

Medidas razonables equivalentes a controles ISO 27001 (sin exigir certificación).

Reporte de incidentes de seguridad en máximo 72 horas.

Transparencia y trazabilidad

Registro básico de versiones y actualizaciones.

Descripción clara de limitaciones y alcances de la herramienta.

Condiciones contractuales

Prohibición de reutilizar información institucional.

Cláusula de confidencialidad reforzada para datos sensibles.

Destrucción de datos una vez finalice el contrato.

4.2.7.4. Matriz simplificada de clasificación de riesgos en casos de uso de GenAI

La siguiente matriz permite clasificar rápidamente el nivel de riesgo del caso de uso:

Tabla 11. Guía de Uso Interno IA: Matriz de Clasificación de Riesgos

Nivel de Riesgo	Características	Requisitos
Bajo	Borradores, textos no sensibles, información pública	Supervisión humana
Medio	Trámites administrativos, datos semisensibles	Validación técnica + supervisor
Alto	Impacto sobre derechos ciudadanos, automatización	Piloto + Aprobación Comité IA
Muy Alto	Datos personales sensibles, decisiones automáticas, impacto jurídico	Prohibido

Fuente: Elaboración Propia

Este cuadro permite que cualquier entidad evalúe rápidamente si un uso está permitido, condicionado o prohibido.

4.2.7.5. Procedimiento de respuesta ante incidentes

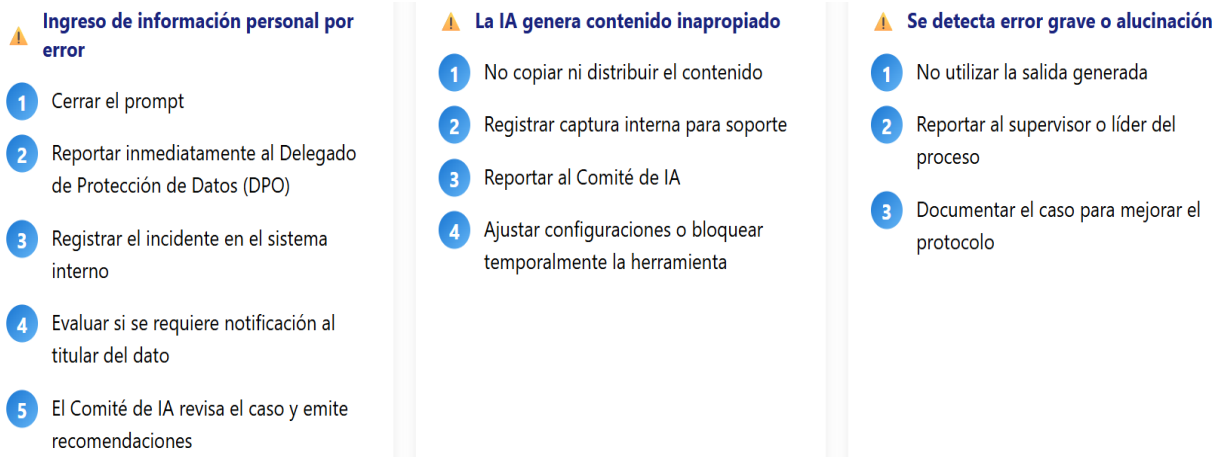
En caso de presentarse un incidente relacionado con el uso de IA generativa, se aplicarán las siguientes acciones según la naturaleza del evento. Si se ingresó información personal por error, se debe cerrar inmediatamente el prompt, reportar el hecho al Delegado de Protección de Datos (DPO), registrar el

incidente en el sistema interno y evaluar la necesidad de notificar al titular del dato. Posteriormente, el Comité de IA revisará el caso y emitirá las recomendaciones correspondientes.

Cuando la IA genere contenido inapropiado, no se debe copiar ni distribuir dicho material. Es necesario registrar una captura interna como evidencia, reportar el incidente al Comité de IA y proceder a ajustar las configuraciones o bloquear temporalmente la herramienta para prevenir recurrencias.

Finalmente, si se detecta un error grave o una alucinación en la salida generada, esta no debe ser utilizada bajo ninguna circunstancia. El hecho debe reportarse al supervisor o líder del proceso y documentarse adecuadamente para fortalecer el protocolo y evitar futuros incidentes. Como se especifica en la Figura 5 se puede identificar el procedimiento de respuesta ante incidentes.

Figura 5 *Guía de Uso Interno: Procedimiento de Respuesta ante Incidentes*



Fuente: Elaboración Propia (2026)

4.2.7.6. Programa obligatorio de capacitación

Para acceder a las herramientas de IA generativa será indispensable completar un programa de formación orientado a garantizar un uso seguro y responsable. Este programa incluye contenidos mínimos obligatorios, impartidos en sesiones presenciales o virtuales sincrónicas, con una duración total de cuatro horas. Los módulos se distribuyen de la siguiente manera: el primero aborda los riesgos asociados a la IA generativa, como alucinaciones, sesgos y aspectos de privacidad, durante una hora; el segundo se centra en la normativa aplicable, incluyendo la Ley 1581 (2012), la Ley 2279 (2022) y las políticas institucionales, también con una hora de duración; el tercero desarrolla la matriz de usos permitidos y prohibidos mediante casos prácticos, en una sesión de hora y media; finalmente, el cuarto módulo explica el protocolo de respuesta ante incidentes en treinta minutos.

Para obtener la certificación será necesario aprobar una evaluación con un puntaje mínimo del 80 %. La vigencia de la certificación será de doce meses, considerando la rápida evolución de la tecnología, y se exigirá una re-certificación cada vez que se produzcan actualizaciones normativas significativas.

La implementación se realizará de manera escalonada. En la primera fase, durante los tres primeros meses, se capacitará a los miembros del Comité de IA, al Delegado de Protección de Datos (DPO) y a los responsables técnicos. La segunda fase, entre los meses cuatro y seis, estará dirigida a funcionarios de áreas estratégicas y misionales. Finalmente, la tercera fase, que abarcará del mes siete al doce, garantizará la cobertura completa para todos los funcionarios usuarios.

El control de cumplimiento se efectuará mediante el sistema de acceso a las herramientas institucionales de IA generativa, que verificará automáticamente la vigencia de la certificación. En caso de no contar con ella, el acceso será suspendido. La responsabilidad de este programa recaerá en el Oficial de IA Responsable, en coordinación con el área de Gestión Humana.

En conjunto, la Guía Institucional de Uso Interno de IA Generativa constituye un instrumento estructural que habilita una adopción responsable, progresiva y segura de estas tecnologías dentro del Distrito Capital. Más que un documento normativo, es un mecanismo de gestión del cambio que articula la protección de derechos, la eficiencia administrativa y la innovación pública. Su diseño flexible permite actualizar fácilmente los lineamientos conforme evolucione la regulación nacional y las capacidades institucionales, consolidando al Distrito como referente nacional en prácticas de IA responsable.

4.3.Análisis de Resultados de Validación

4.3.1. Selección de Caso de Simulación

Para la validación experimental del framework se seleccionó el caso de uso del Sistema Automatizado de Validación y Expedición de Certificados de Residencia. La elección de este caso, en lugar del de Objeciones de Comparendos, se fundamentó en su mayor complejidad y nivel de riesgo inherente. El sistema de certificados de residencia fue clasificado como de Alto Riesgo debido a que toma decisiones que impactan directamente el acceso de los ciudadanos a un servicio esencial, lo que a su vez puede afectar otros derechos y beneficios. Esta clasificación activa la totalidad de las fases y controles del framework, incluyendo la obligatoriedad de un Análisis de Riesgos Algorítmicos y Evaluación de Impacto en Protección de Datos (ARA/DPIA) exhaustivo. Por el contrario, el caso de objeciones fue clasificado como de Riesgo Limitado, ya que su función principal es de asistencia y clasificación preliminar, sin tomar una decisión final sobre el ciudadano. Por lo tanto, el caso del Certificado de Residencia representa un escenario más retador y completo, generando mayor valor demostrativo

para simular y validar la robustez y aplicabilidad de las nueve fases del ciclo de vida de gobernanza propuesto.

4.3.2. Validación Experimental mediante Simulación

La validación experimental documenta la aplicación del framework cada una de las 9 fases del ciclo de vida de gobernanza de IA detallando cómo se ha llevado a cabo cada una de las actividades, se aplica el modelo de gobierno con la identificación los diferentes roles y responsabilidades de cada fase, se explica el uso de las herramientas y se identifican los correspondientes entregables que se usarán de evidencia hacia los Gates llevados a cabo por el comité de IA. Para estandarizar la documentación de cada fase se generó una guía de implementación del ciclo de vida de gobernanza de IA disponible en el Anexo D, completamente funcional y que detalla la interrelación con todos los componentes documentados en desarrollo de la propuesta de framework y aplicada como base y lienzo sobre todo el ejercicio de simulación.

4.3.2.1. Simulación de Ejecución: Fase 1 - Intake y AI Use-Case Canvas

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 1 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito y basado en la información del AI Use-Case Canvas.

Contexto de la Iniciativa

Entidad: Secretaría de Gobierno del Distrito Capital. Sponsor de Negocio: Director de Atención al Ciudadano / Product Owner. Responsable Técnico: Líder de Desarrollo e Innovación / Equipo TIC. Fecha de Solicitud: 26/11/2025 - v1.0.

Actividad 1: Completar AI Use-Case Canvas

El Sponsor de Negocio ha liderado la creación de la propuesta, documentando las dimensiones clave del proyecto en la herramienta AI Use-Case Canvas visible en el Anexo E1.

A. Definición del Problema y Objetivos Problema a Resolver: El proceso actual implica una validación manual de documentos que genera tiempos de espera de hasta 24 horas, depende de la capacidad humana limitada y horarios de oficina, y presenta riesgos de error en la subsanación. Propósito del Sistema: Automatizar la clasificación y validación de documentos (Cédula y Recibos) mediante IA para emitir el certificado de forma inmediata y disponible 24/7. Métricas de Éxito Esperadas: Reducción del 95% en el tiempo de expedición (de 24h a minutos). Aumento del 23% en la satisfacción ciudadana (CSAT). Disponibilidad del servicio 24/7.

B. Actores y Datos Usuarios: Ciudadanos solicitantes del certificado (población general de Bogotá). Perfil: Heterogéneo, con niveles variados de alfabetización digital y acceso a dispositivos de diferente calidad. Datos Requeridos: Imágenes o PDFs cargados por el ciudadano (Documento de Identidad, Recibo de Servicio Público). Categoría de Datos: Personales (nombres, direcciones, número de documento de identidad).

C. Identificación Preliminar de Riesgos Se han identificado riesgos tempranos que activan principios de gobernanza: Equidad: Riesgo de sesgo técnico (OCR) que falle más con documentos de baja calidad, discriminando a ciudadanos con menor acceso a tecnología. Privacidad: Exposición de datos personales si la seguridad en el manejo de los archivos temporales es débil. Operativos: Sobrecarga del personal humano si la tasa de derivación de casos es muy alta.

Actividad 2: Valoración de Viabilidad

El borrador del canvas fue sometido a un "filtro de viabilidad" por los roles clave:

Responsable Técnico: Confirma la viabilidad técnica de la solución (OCR/NLP) y establece KPIs técnicos (Precisión ≥ 98%). Área Jurídica: Valida la base legal (Ejercicio de funciones públicas y simplificación de trámites) y la conformidad con la Ley 1581 (2012).

Actividad 3: Identificación de Riesgos y Partes Interesadas

Se realizó una identificación temprana de riesgos y actores clave.

Delegado de Protección de Datos (DPO): Identifica la necesidad de un ARA/DPIA en fases posteriores debido al tratamiento masivo de datos personales y la toma de decisiones automatizada. Stakeholders: Se identificaron a los ciudadanos, funcionarios de validación y el equipo de TI como los principales actores.

Punto de Control (Gate 1): Revisión y Decisión del Comité de IA

El Comité de IA evaluó la propuesta considerando la alineación estratégica, viabilidad y claridad del propósito.

Decisión Final

ESTADO: APROBADO

Instrucción: Se aprueba el paso a la Fase 2 - Clasificación de Riesgo.

Observaciones del Comité: Se condiciona el desarrollo a la presentación de un plan de mitigación de sesgos por calidad de imagen. Se debe implementar un flujo de 'Human-in-the-loop' para casos de baja confianza, asegurando que no se rechacen automáticamente.

4.3.2.2. Simulación de Ejecución: Fase 2 - Clasificación de Riesgo

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 2 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase se activa tras la aprobación del AI Use-Case Canvas en el Gate 1.

Contexto de la Evaluación

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Sesión: 28/11/2025. Participantes (Matriz RACI): Accountable (A): Comité de IA. Responsible (R): Responsable Técnico (Líder de Desarrollo e Innovación). Consulted (C): DPO y Área Jurídica. Informed (I): Sponsor de Negocio.

Actividad 1: Evaluación de Matriz de Clasificación

El equipo del proyecto, liderado por el Responsable Técnico, registró la Matriz de Clasificación de Riesgo documentada en Anexo E2, evaluó el caso de uso contra los criterios definidos en la Política de Gestión de Riesgos de IA, alineada con el AI Act y CONPES 4144 (2024).

Análisis de Criterios

1. Propósito del Sistema: Validar documentos (Cédula y Recibos) para la expedición de un acto administrativo (Certificado de Residencia). Evaluación: El sistema actúa como un filtro de acceso a un servicio público esencial.
2. Población Afectada y Tipo de Decisión: Afecta a la ciudadanía general de Bogotá. La decisión (aprobar/rechazar) impacta el acceso a un derecho fundamental y es prerrequisito para otros trámites y subsidios. Evaluación: Impacto significativo en derechos y servicios esenciales.
3. Datos Procesados: Datos personales (Nombres, ID, Direcciones). No utiliza datos biométricos para identificación remota en espacios públicos (solo validación documental 1:1). Evaluación: Tratamiento masivo de datos personales.
4. Consecuencias de Errores: Un falso negativo (rechazo incorrecto) genera barreras administrativas injustificadas y vulnera el debido proceso. Evaluación: Impacto alto en el individuo, aunque reversible mediante intervención humana.

Resultado de la Clasificación

NIVEL DE RIESGO PRELIMINAR: ALTO RIESGO

Actividad 2: Asignación de Nivel de Riesgo

Justificación de la Clasificación

El sistema se clasifica como Alto Riesgo porque cumple con las siguientes condiciones de la taxonomía: Impacto en Servicios Esenciales: Interviene en la provisión de un servicio público esencial. Decisión Automatizada: Asiste en decisiones que tienen efectos jurídicos sobre las personas.

¿Por qué NO es Riesgo Inaceptable? Se verificó que el sistema: NO realiza puntuación social (Social Scoring). NO utiliza técnicas subliminales. NO realiza identificación biométrica remota masiva en tiempo real. NO explota vulnerabilidades de grupos específicos.

Actividad 3: Activación de Obligaciones Reforzadas

Dada la clasificación de Alto Riesgo, se activan automáticamente las siguientes obligaciones reforzadas para las fases subsiguientes:

1. Fase 3 (ARA/DPIA): Es OBLIGATORIO realizar una Evaluación de Impacto en Protección de Datos y Riesgos Algorítmicos completa.
2. Supervisión Humana: Se exige implementar un mecanismo de Human-in-the-loop (HITL). No se permiten rechazos automáticos sin revisión.
3. Documentación Técnica: Se requerirá una Model Card detallada y Data Sheets exhaustivos (Fases 4 y 6).
4. Pruebas de Equidad: En la Fase 6, será obligatorio demostrar que la tasa de error no varía significativamente entre documentos digitales y físicos (riesgo de sesgo por calidad de imagen).
5. Auditoría: El sistema estará sujeto a auditorías externas anuales en la Fase 8.

Punto de Control (Gate 2): Validación de la Clasificación

El Comité de IA revisó la propuesta de clasificación y la justificación presentada.

Decisión Final del Comité

DECISIÓN: VALIDADA

Dictamen: El Comité de IA ratifica la clasificación de Alto Riesgo.

Instrucción al Equipo: Proceder inmediatamente a la Fase 3 - ARA/DPIA involucrando al DPO. El proyecto NO puede avanzar a desarrollo (Fase 5) hasta que el ARA/DPIA sea aprobado en el Gate 3.

4.3.2.3. Simulación de Ejecución: Fase 3 - ARA/DPIA (Alto Riesgo)

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 3 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase es obligatoria dado que el sistema fue clasificado como de Alto Riesgo en la Fase 2.

Contexto de la Evaluación

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Sesión: 02/12/2025. Participantes (Matriz RACI): Accountable (A): Comité de IA y DPO (Voto Vinculante). Consulted (C): Responsable Técnico y Sponsor de Negocio.

Actividad 1: Completar Plantilla ARA/DPIA

El DPO, junto con el Responsable Técnico, completó la Plantilla ARA/DPIA, documentada en Anexo E3, para evaluar los impactos en derechos fundamentales y privacidad.

A. Mapeo de Datos y Base Legal Datos Tratados: Imágenes de Cédulas de Ciudadanía y Recibos de Servicios Públicos. Categoría: Datos Personales (Ley 1581, 2012). No son datos sensibles biométricos (no se hace reconocimiento facial 1:N), pero sí datos de alto impacto administrativo. Base Legal: Cumplimiento de una obligación legal y ejercicio de funciones públicas (Ley 2052 de 2020 - Simplificación de trámites). Flujo: Carga por ciudadano -> Procesamiento en memoria (OCR) -> Validación -> Eliminación de archivo fuente -> Generación de Certificado.

Actividad 2: Análisis de Riesgos

B. Evaluación de Impactos en Derechos (Hallazgos Clave)

1. Derecho a la Igualdad (Equidad): Riesgo Identificado: El modelo OCR podría tener una tasa de error mayor con fotos de baja calidad (celulares gama baja), discriminando indirectamente a poblaciones vulnerables. Nivel: Alto.
2. Debido Proceso: Riesgo Identificado: Un "falso negativo" (rechazo erróneo) automático impediría el acceso a un derecho sin intervención humana. Nivel: Alto.
3. Privacidad: Riesgo Identificado: Exposición de datos si los archivos temporales no se eliminan o si el proveedor los usa para entrenar sus propios modelos. Nivel: Medio.

Actividad 3: Propuestas de Medidas de Mitigación

Para gestionar los riesgos identificados, se definieron las siguientes medidas de mitigación obligatorias que el equipo técnico debe implementar en la Fase 5 (Desarrollo):

Medidas Técnicas

1. Protocolo "Human-in-the-loop" (HITL): Se prohíbe el rechazo automático. Si la confianza del modelo es <90% o la imagen es borrosa, el caso se deriva a una bandeja de revisión humana.
2. Privacidad por Diseño: Configuración de eliminación automática (TTL) de las imágenes cargadas inmediatamente después de la validación. Cifrado en tránsito y reposo.

Medidas Organizativas

1. Certificación de Datos: El Propietario de Datos debe garantizar (en Fase 4) que el dataset de entrenamiento incluye una muestra representativa de recibos físicos, arrugados y de todos los proveedores de servicios de Bogotá.
2. Prohibición de Usos Secundarios: Se prohíbe explícitamente usar los datos extraídos (estrato, consumo) para scoring crediticio o mapas de calor de morosidad.

Medidas de Validación (Criterios de Aceptación para Fase 6)

1. Prueba de Equidad: La diferencia en la tasa de error entre documentos de alta calidad (digitales) y baja calidad (fotos) no debe superar el 5%.

Punto de Control (Gate 3): Aprobación Vinculante del DPO

El documento ARA/DPIA y el plan de mitigación fueron presentados al Comité de IA.

Decisión del DPO (Voto Vinculante)

DECISIÓN: APROBADO CON CONDICIONES

Dictamen del DPO: "Apruebo el análisis de impacto bajo la condición estricta de que se implementen las cláusulas de confidencialidad reforzadas con el proveedor tecnológico y se verifique que no se utilicen los datos de los ciudadanos para re-entrenar modelos externos. La medida de Human-in-the-loop es indispensable para mitigar el riesgo de debido proceso."

Ratificación del Comité de IA

ESTADO: APROBADO. Se autoriza el paso a la Fase 4 - Gobierno de Datos.

Instrucción: El plan de mitigación se convierte en requisitos funcionales obligatorios para el desarrollo.

4.3.2.4. Simulación de Ejecución: Fase 4 - Gobierno de Datos

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 4 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase asegura que los datos utilizados para entrenar el modelo sean de alta calidad, representativos y gestionados éticamente.

Contexto de la Ejecución

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Sesión: 05/12/2025. Participantes (Matriz RACI): Accountable (A): DPO (Oficial de Protección de Datos). Responsible (R): Responsable Técnico y Propietario de Datos (Data Steward). Consulted (C): Comité de IA y Sponsor de Negocio.

Actividad 1: Inventario y Documentación de Fuentes

El Propietario de Datos (Data Steward), junto con el equipo técnico, realizó una auditoría exhaustiva del conjunto de datos denominado RESIDENCIA_BOG_VALIDATION_V1.

Fuentes: Datos históricos de solicitudes previas anonimizadas y cargas controladas durante la fase piloto.

Actividad 2: Evaluación de Calidad Dimensional

Volumen: 15,000 registros (Imágenes PDF/JPG y texto extraído). Completitud: Menos del 2% de datos faltantes en campos no estructurados. Consistencia: Se identificaron inconsistencias en formatos de dirección (ej. "Cll" vs "Calle"), las cuales fueron normalizadas.

Actividad 3: Análisis de Representatividad y Detección de Sesgos

Durante el análisis estadístico, se detectó un riesgo significativo de sesgo en la composición original del dataset: Hallazgo: El 80% de las facturas correspondían al proveedor Enel en formato nativo digital. Riesgo: El modelo podría fallar sistemáticamente con recibos de otros proveedores (Acueducto, Vanti) o con formatos físicos escaneados, discriminando a ciudadanos que no reciben factura digital. Acción Correctiva: Se aplicó una técnica de balanceo (oversampling) enriqueciendo el dataset con muestras de otros proveedores y fotografías de baja calidad tomadas con celulares de gama baja.

Actividad 4: Elaboración de Data Sheets

Se documentó el conjunto de datos utilizando la plantilla estándar de Data Sheet del framework, documentada en el Anexo E4, asegurando transparencia sobre su origen y limitaciones.

Resumen del Data Sheet (RESIDENCIA_BOG_VALIDATION_V1) Propósito: Entrenar modelos OCR/NLP para validación de identidad y residencia. Datos Personales: Sí (Nombres, Cédulas, Direcciones). Datos

Sensibles: No. Base Legal: Ejercicio de funciones públicas y simplificación de trámites. Transformaciones: Pre-procesamiento de imágenes (contraste) y normalización de texto. Usos Prohibidos: Se estableció explícitamente la prohibición de usar estos datos para evaluar capacidad de pago, crear mapas de morosidad o compartir con terceros comerciales.

Punto de Control (Gate 4): Certificación de la Calidad de los Datos

El Propietario de Datos y el DPO realizaron la revisión final del Data Sheet y la calidad del dataset enriquecido.

Certificación del Propietario de Datos y DPO

DECISIÓN: APROBADO CON CONDICIÓN DE EQUIDAD

Dictamen: "Se certifica que el dataset cumple con los estándares de calidad y que se han aplicado las medidas correctivas para mitigar el sesgo de proveedor.

Condición para Fase 6: No se autoriza el despliegue hasta que se demuestre estadísticamente en la Fase de Pruebas que el modelo reconoce con igual precisión (diferencia <5%) una factura digital y una fotografía de celular de un recibo físico."

Instrucción: El dataset queda habilitado para ser utilizado por el equipo de desarrollo en la Fase 5.

4.3.2.5. Simulación de Ejecución: Fase 5 - Desarrollo y Adquisición

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 5 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase materializa los requisitos éticos y legales definidos en las fases anteriores (Canvas, Riesgo, ARA/DPIA) en obligaciones contractuales y especificaciones técnicas tangibles.

Contexto de la Ejecución

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Cierre de Fase: 15/12/2025. Estrategia de Implementación: Híbrida (Adquisición de Motor OCR + Desarrollo Interno de Integración). Participantes (Matriz RACI): Accountable (A): Responsable Técnico (Desarrollo) y Área Jurídica (Contratación). Consulted (C): DPO y Comité de IA.

Camino B: Adquisición Externa

Actividad 1: Debida Diligencia y Selección del Proveedor (Motor OCR)

Dado que la entidad no cuenta con la capacidad para desarrollar un motor de OCR/NLP desde cero, se procedió a la adquisición de una solución de mercado. Se aplicó el Checklist de Evaluación de Proveedores de IA (versión Estándar para Alto Riesgo), documentado en el Anexo E5.

Proveedor Evaluado: VisionTech OCR Services.

Resultado del Checklist: RECOMENDADO (Cumplimiento de criterios mandatorios).

Hallazgos Clave: Conformidad Regulatoria: El proveedor presentó certificación de conformidad con la Ley 1581 (2012) y aceptó firmar el DPA del Distrito. Transparencia (Caja Blanca): Se obligó contractualmente a la entrega de la Model Card del modelo base y el Data Sheet de sus datos de entrenamiento. Seguridad: Cuenta con certificación ISO 27001 vigente. Auditoría: Aceptó la cláusula de "Derecho a Auditoría" por parte del Distrito.

Actividad 2: Negociación del Contrato con Cláusulas de Gobernanza

El Área Jurídica, con apoyo del DPO, blindó la contratación del motor OCR mediante cláusulas específicas.

Acuerdo de Procesamiento de Datos (DPA): Define al Distrito como Responsable y a VisionTech como Encargado. Prohíbe el uso de los datos del Distrito para re-entrenar los modelos comerciales del proveedor. SLA de Precisión: Se establece un nivel de servicio donde la precisión del OCR no puede ser inferior al 95%. Penalizaciones económicas por degradación del modelo. Portabilidad: Obligación de devolver o destruir todos los datos al finalizar el contrato.

Camino A: Desarrollo Interno

Actividad 1: Diseño e Implementación con Gobernanza Integrada

El equipo de desarrollo interno de la Secretaría construyó la capa de orquestación y la interfaz de usuario (Chatbot), implementando los controles definidos en el ARA/DPIA (Fase 3).

Implementación de Controles en Código

1. Protocolo "Human-in-the-loop" (HITL): Implementación: Se desarrolló un microservicio de "Gestión de Excepciones". Lógica: Si confianza_ocr < 0.90 OR calidad_imagen == 'baja', el sistema no emite respuesta de rechazo. En su lugar, encola la solicitud en el dashboard de los funcionarios supervisores.

2. Privacidad por Diseño (Minimización): Implementación: Configuración de políticas de ciclo de vida en el almacenamiento (Bucket S3). Regla: Las imágenes de cédulas y recibos tienen un TTL (Time-to-Live) de 1 hora tras la finalización del trámite. Solo se persisten los metadatos y logs de auditoría.
3. Trazabilidad Inmutable: Implementación: Sistema de logs centralizado que registra cada decisión: Input_ID, Score_Confianza, Decisión_IA, Decisión_Humana (si aplica), Timestamp.

Punto de Control (Gate 5): Revisión de la Solución y Contrato

El Comité Técnico y Jurídico revisó los entregables antes de autorizar el paso a pruebas.

Decisión

DECISIÓN: APROBADO

Dictamen: "Se aprueba la arquitectura híbrida y el contrato con VisionTech OCR Services. Verificación Técnica: La integración implementa correctamente el flujo de derivación a humanos (HITL). Verificación Jurídica: El contrato incluye las salvaguardas de la Ley 1581 (2012) y garantías de auditoría.

Instrucción: Proceder a la Fase 6 - Pruebas y Validación, donde se deberá ejecutar el test de equidad (diferencia de error < 5%) exigido en la Fase 3."

4.3.2.6. Simulación de Ejecución: Fase 6 - Pruebas y Validación

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 6 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase tiene como objetivo demostrar con evidencia empírica que el sistema cumple con los requisitos técnicos, éticos y funcionales definidos en las fases previas (especialmente en el ARA/DPIA).

Contexto de la Ejecución

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Cierre de Fase: 20/01/2026. Participantes (Matriz RACI): Accountable (A): Responsable Técnico (Líder de Desarrollo). Consulted (C): Comité de IA, DPO. Informed (I): Área Jurídica.

Actividad 1: Pruebas Técnicas

El equipo técnico ejecutó una batería de pruebas sobre el modelo RESIDENCIA_BOG_OCR_VALIDATOR_V1.0 utilizando el conjunto de datos de prueba reservado (10% del dataset total, nunca visto por el modelo).

A. Rendimiento del Modelo Se evaluaron las métricas de desempeño contra los KPIs definidos en la Fase 1: Precisión Global (Accuracy): 96.5% (Meta: ≥ 95%). CUMPLE Tasa de Resolución Autónoma: 92% de los documentos fueron procesados con una confianza superior al 90%. CUMPLE Tiempo de Respuesta: Promedio de 1.8 segundos por documento. CUMPLE

B. Robustez Técnica Se sometió al modelo a pruebas de estrés con imágenes degradadas: Prueba de Ruido: Se inyectó ruido gaussiano a las imágenes. El modelo mantuvo una precisión >90% hasta niveles de ruido moderado. Prueba de Rotación: El modelo corrigió automáticamente rotaciones de hasta 45 grados.

C. Seguridad (Ciberseguridad) Se realizó un pentesting enfocado en riesgos de IA: Ataques de Evasión: Se intentó engañar al OCR modificando píxeles imperceptibles. El modelo mostró resistencia adecuada. Inyección de Prompts (Chatbot): Se verificó que el componente de chat no revelara instrucciones del sistema ni datos de otros usuarios.

Actividad 2: Pruebas de Equidad

Esta fue la prueba crítica condicionada por el DPO y el Comité de IA en la Fase 3 y 4.

Objetivo: Verificar que el sistema no discrimine a ciudadanos que aportan documentos físicos (fotos de celular) frente a los que aportan documentos digitales (PDFs originales).

Resultados Cuantitativos: Tasa de Acierto en Documentos Digitales (Alta Calidad): 98.5% Tasa de Acierto en Fotos de Celular (Baja/Media Calidad): 94.7% Diferencia (Gap de Equidad): 3.8%

Evaluación Criterio de Aceptación: La diferencia no debe superar el 5%. Resultado: 3.8% < 5%. CUMPLE

Observación: Aunque cumple el umbral, la diferencia existente justifica plenamente la medida de mitigación de "Human-in-the-loop" para los casos que caen en ese margen de error.

Actividad 3: Pruebas de Explicabilidad

Se validó que el sistema sea transparente para el usuario final y explicable para el auditor.

Notificación al Usuario: Se verificó que el Chatbot inicia la conversación con el mensaje: "Hola, soy un asistente virtual automatizado de la Secretaría de Gobierno...". Explicabilidad de la Decisión: En caso de aprobación: El sistema informa qué datos validó. En caso de derivación a humano: El sistema informa "La calidad de la imagen no permite una validación automática segura. Un funcionario revisará su caso en breve". No se generan rechazos "caja negra".

Actividad 4: Pruebas de Usabilidad

Accesibilidad: Se auditó la interfaz web del Chatbot cumpliendo con WCAG 2.1 Nivel AA (compatible con lectores de pantalla). Usabilidad: Se realizaron pruebas con ciudadanos de diferentes perfiles para asegurar que la interacción con el chatbot fuera intuitiva.

Actividad 5: Pruebas de Integración

Integración: Se verificó la correcta comunicación con la base de datos de radicación y la generación del PDF del certificado firmado digitalmente. Prueba de Carga: El sistema soportó 500 peticiones concurrentes sin degradación del servicio (simulando picos de demanda).

Punto de Control (Gate 6): Aprobación para Despliegue

El Responsable Técnico presentó el Informe de Pruebas y Validación y la Model Card actualizada al Comité de IA, y que se puede ver documentada en el Anexo E6.

Decisión del Comité de IA

DECISIÓN: APROBADO PARA DESPLIEGUE (GO-LIVE)

Dictamen: "La evidencia presentada demuestra que el sistema es técnicamente robusto y cumple con los criterios de equidad establecidos en el ARA/DPIA.

Instrucciones para Fase 7 (Despliegue):

1. Proceder con el despliegue gradual (Piloto en 2 localidades) durante las primeras 2 semanas.
2. Activar el monitoreo intensivo de 'Drift' para asegurar que los datos de producción se comporten igual que los de prueba."

4.3.2.7. Simulación de Ejecución: Fase 7 - Despliegue

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 7 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase marca la transición crítica del entorno de pruebas al entorno de producción, asegurando una puesta en marcha controlada.

Contexto de la Ejecución

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Sesión (Gate 7): 25/01/2026. Participantes (Matriz RACI): Accountable (A): Comité de IA (Autorización Final). Responsible (R): Responsable Técnico. Consulted (C): Sponsor de Negocio y DPO.

Actividad 1: Capacitación de Usuarios

Antes de activar el sistema, se ejecutó el plan de formación para garantizar que el componente humano del sistema ("Human-in-the-loop") fuera competente.

Funcionarios Supervisores: Se capacitó al equipo de Atención al Ciudadano en el manejo de la "Bandeja de Excepciones". Enfoque: Cambio de rol de "validadores de todo" a "gestores de excepciones". Protocolo: Instrucción clara de que ante la duda razonable en una imagen borrosa, prevalece el criterio humano garantista (pro-ciudadano). Ciudadanía: Se publicaron infografías en el portal web explicando el nuevo trámite inmediato y cómo tomar correctamente la foto del recibo para evitar rechazos técnicos.

Actividad 2: Configuración de Controles

El equipo técnico activó la infraestructura de observabilidad definida en el ARA/DPIA:

Logs de Auditoría: Activación del registro inmutable de transacciones (Input_ID, Score_Confianza, Decisión, Timestamp). Dashboard de Gobernanza: Se habilitó el tablero en tiempo real para monitorear: Tasa de Derivación a Humanos (Meta: <15%). Tiempos de Respuesta (Meta: <3 min). Alertas de Drift (si la distribución de proveedores de recibos cambia drásticamente).

Actividad 3: Despliegue Gradual

Siguiendo la instrucción del Gate 6, se optó por un lanzamiento faseado para minimizar riesgos operativos.

Fase A (Semana 1-2): Piloto Controlado. Alcance: Habilitado solo para solicitudes originadas en las localidades de Kennedy y Suba. Objetivo: Validar carga y comportamiento con documentos reales en zonas de alta demanda. Fase B (Semana 3+): Escalamiento General. Condición: Si no se presentan incidentes críticos en Fase A, se abre a todo el Distrito.

Actividad 4: Establecimiento de Canales de Reporte

Se implementaron los mecanismos de transparencia y defensa del ciudadano:

Botón de Apelación: En caso de que el sistema (o el humano supervisor) rechace la solicitud, se habilitó un botón visible: "No estoy de acuerdo, solicitar segunda revisión", que escala el caso a un nivel superior.

Actividad 5: Comunicación Transparente

Transparencia Activa: El Chatbot inicia con el mensaje: "Hola, soy un asistente virtual automatizado. Estoy aquí para validar tus documentos y expedir tu certificado al instante."

Punto de Control (Gate 7): Aprobación del Go-Live

El Comité de IA se reunió con el Sponsor de Negocio para la autorización final.

Verificación de Pre-requisitos

1. ¿Personal capacitado? SÍ.
2. ¿Monitoreo activo? SÍ.
3. ¿Plan de comunicación ejecutado? SÍ.
4. ¿Pruebas de equidad superadas (Fase 6)? SÍ.

Decisión Final

DECISIÓN: AUTORIZADO EL GO-LIVE (Fase A - Piloto)

Dictamen: "Se autoriza la puesta en producción del sistema bajo la modalidad de piloto controlado en Kennedy y Suba.

Condición de Operación: Se declara estado de Hypercare (monitoreo intensivo) durante las próximas 2 semanas. El Responsable Técnico debe reportar diariamente al Comité de IA cualquier anomalía en la tasa de rechazos."

4.3.2.8. Simulación de Ejecución: Fase 8 - Monitoreo y Auditoría

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 8 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase abarca el primer trimestre de operación tras el despliegue, enfocándose en la vigilancia continua y la rendición de cuentas.

Contexto de la Ejecución

Entidad: Secretaría de Gobierno del Distrito Capital. Periodo Evaluado: Primer Trimestre de Operación (Febrero - Abril 2026). Fecha de Sesión (Gate 8): 30/04/2026. Participantes (Matriz RACI): Accountable (A): Comité de IA. Responsible (R): Responsable Técnico. Consulted (C): DPO y Sponsor de Negocio.

Actividad 1: Monitoreo Técnico

El Responsable Técnico presentó el reporte del Dashboard de Gobernanza con los datos acumulados de los primeros 90 días de operación.

A. Métricas de Desempeño y Servicio Volumen: 45,000 solicitudes procesadas. Tasa de Resolución Autónoma: 91.5% (Meta: >90%). CUMPLE Observación: El sistema validó automáticamente la gran

mayoría de casos, liberando carga operativa. Tiempo Promedio de Respuesta: 1.5 minutos (Meta: <3 min). CUMPLE Satisfacción Ciudadana (CSAT): 4.6/5.0 (Meta: >4.2). CUMPLE

B. Métricas de Equidad y Drift Monitoreo de Equidad: Tasa de error en documentos digitales: 1.2%. Tasa de error en fotos de celular (gama baja): 4.5%. Gap: 3.3% (Meta: <5%). CUMPLE Detección de Drift: Se detectó una leve desviación en la segunda semana de marzo debido a un cambio estético en la factura de un proveedor menor de internet (no contemplado inicialmente). Acción: El sistema derivó estos casos a humanos (baja confianza) correctamente. No se requirió reentrenamiento urgente, pero se agendó para el próximo ciclo.

Actividad 2: Gestión de Incidentes

Se revisó el registro de incidentes reportados durante el trimestre.

Incidente Destacado: INC-2026-003 (Queja por Rechazo) Descripción: Un ciudadano reportó vía PQR que el sistema rechazó su certificado alegando "Dirección no coincide", aunque él afirmaba que sí. Investigación: Se revisaron los logs de auditoría. La IA marcó confianza baja (85%) por una abreviatura no estándar en la dirección. El caso fue derivado a un supervisor humano (protocolo HITL). El supervisor humano rechazó la solicitud por error de lectura. Resolución: Se contactó al ciudadano, se corrigió el error manualmente y se expidió el certificado. Lección Aprendida: Se reforzó la capacitación de los supervisores humanos sobre abreviaturas no estándar. El caso se etiquetó para futuro reentrenamiento (Golden Dataset).

Actividad 3: Auditoría Periódica (Interna)

El equipo de Control Interno realizó una verificación muestral de cumplimiento.

Verificación de Supervisión Humana: Se auditaron 50 casos derivados a la "Bandeja de Excepciones". En el 100% de los casos hubo una acción registrada por un funcionario (Aprobar/Rechazar) con su respectiva justificación. Verificación de Privacidad (Retención de Datos): Se verificó aleatoriamente si existían imágenes de cédulas de solicitudes cerradas en febrero. Resultado: No se encontraron archivos. El script de borrado seguro (TTL 30 días) funcionó correctamente. CUMPLE

Actividad 4: Gestión de Cambios

Evaluación de Impacto: El drift detectado en marzo fue analizado. Se determinó que no afectaba la precisión de forma crítica, pero se agendó una actualización. Actualización de Documentación: Se creó una nueva versión del Data Sheet para incluir los nuevos formatos de factura y se planificó el reentrenamiento del modelo para la versión v1.1.

Punto de Control (Gate 8): Revisión Trimestral de Desempeño

El Comité de IA analizó la evidencia presentada para decidir el futuro del sistema.

Evaluación del Comité

1. Desempeño: El sistema supera las metas de eficiencia y satisfacción.
2. Riesgos: Los riesgos de equidad están controlados dentro de los márgenes aceptables.
3. Cumplimiento: La auditoría confirma la adherencia a la política de datos.

Decisión Final

DECISIÓN: MANTENER

Dictamen: "El sistema opera correctamente y genera valor público. Se autoriza la continuidad de la operación.

Instrucciones para el próximo trimestre:

1. Recopilar los casos de facturas con nuevos formatos (detectados por drift) para preparar un reentrenamiento programado (v1.1) en el siguiente ciclo.
2. Mantener la vigilancia sobre la tasa de error en fotos de baja calidad."

4.3.2.9. Simulación de Ejecución: Fase 9 - Retiro o Fin de Vida

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia

Este documento simula la ejecución de la Fase 9 del Ciclo de Vida de Gobernanza de IA, siguiendo los lineamientos del Framework de Gobernanza de IA del Distrito. Esta fase cierra el ciclo de vida del sistema, asegurando un desmantelamiento seguro, legal y ordenado.

Contexto de la Ejecución

Entidad: Secretaría de Gobierno del Distrito Capital. Fecha de Sesión (Gate 9): [Fecha Futura Simulada, ej. 30/11/2028]. Participantes (Matriz RACI): Accountable (A): Comité de IA y DPO (Voto Decisivo en Datos). Responsible (R): Responsable Técnico. Consulted (C): Sponsor de Negocio.

Actividad 1: Decisión de Retiro

El Comité de IA ha decidido retirar el sistema basándose en uno de los triggers definidos en la Fase 1.

Causal de Retiro (Simulada) Obsolescencia Técnica: La tecnología OCR utilizada (v1.0) ha sido superada por nuevos modelos de IA Generativa Multimodal que ofrecen mayor precisión a menor costo, haciendo insostenible el mantenimiento del actual.

Actividad 2: Planificación de la Transición

Plan de Transición (Continuidad del Servicio) Estrategia: Migración a nuevo sistema (v2.0). Acción: Se notifica a la ciudadanía con 30 días de antelación mediante el portal web y el mismo chatbot. Mensaje: "Este asistente dejará de funcionar el día [Fecha]. A partir de entonces, el trámite se realizará a través de la nueva plataforma Distrital de Servicios."

Actividad 3: Gestión de Datos

El Propietario de Datos y el DPO supervisaron la disposición final de los activos de información.

Acciones Ejecutadas

1. Eliminación Segura (Secure Wipe): El Responsable Técnico ejecutó scripts de borrado seguro (sobreescritura) de todas las imágenes de cédulas y recibos almacenados en "hot storage" y copias de seguridad temporales.
2. Archivado Legal (Cold Storage): Se conservaron únicamente los logs de auditoría y los metadatos de las transacciones (ID de radicado, fecha, resultado de validación) durante el tiempo exigido por la Ley de Archivo General de la Nación (5 años) para responder a futuras acciones legales.
3. Destrucción del Modelo: El modelo neuronal entrenado fue archivado como activo de propiedad intelectual del Distrito, pero desconectado de los entornos de producción.

Actividad 4: Documentación de Lecciones Aprendidas

Se realizó una sesión de cierre ("Post-Mortem") para documentar el conocimiento adquirido.

Informe de Lecciones Aprendidas Lo que funcionó: La integración con la base de datos de servicios públicos redujo los tiempos en un 95%. El protocolo "Human-in-the-loop" evitó crisis reputacionales por falsos negativos. Lo que falló: El OCR tuvo dificultades persistentes con recibos muy arrugados en zonas rurales, requiriendo más intervención humana de la planeada inicialmente. Recomendación: Para la v2.0, utilizar modelos multimodales que entiendan mejor el contexto visual de documentos deteriorados.

Punto de Control (Gate 9): Aprobación Final y Cierre

El Comité de IA, el DPO y Auditoría Interna se reunieron para el cierre formal.

Verificación de Requisitos (Checklist de Cierre)

1. ¿Se ha garantizado la continuidad del servicio por otro canal? SÍ.

2. (Veto del DPO) ¿Existe certificado técnico de eliminación segura de las imágenes de documentos de identidad? SÍ.
3. ¿Se han revocado los accesos y claves API del proveedor externo? SÍ.
4. ¿Está guardado el informe de Lecciones Aprendidas? SÍ.

Decisión Final

DECISIÓN: APROBADO EL RETIRO DEFINITIVO

Dictamen: "Se declara la terminación del ciclo de vida del sistema 'Chatbot Certificado Residencia v1.0'. Se instruye a TI liberar los recursos de nube asociados y archivar el expediente en el Registro Central de Casos de Uso de IA."

Documento generado para el cierre del expediente del proyecto en el Registro Central de Casos de Uso de IA.

4.3.3. Hallazgos Emergentes y Ajustes Finales

La validación experimental mediante la simulación del caso de uso del Certificado de Residencia a través de las nueve fases del framework permitió identificar hallazgos clave y realizar ajustes finales que robustecen el modelo de gobernanza.

4.3.3.1. Hallazgos Emergentes:

1. Criticidad de la Clasificación de Riesgo (Fase 2): La simulación confirmó que la clasificación inicial como 'Alto Riesgo' fue la decisión más determinante del ciclo de vida, ya que activó correctamente las obligaciones reforzadas (ARA/DPIA, pruebas de equidad, auditoría externa) que resultaron ser indispensables para mitigar los riesgos más significativos.
2. Efectividad del ARA/DPIA (Fase 3): El análisis de impacto no fue un ejercicio teórico. Permitió identificar riesgos concretos y no evidentes a primera vista, como el sesgo por calidad de imagen en el OCR, y transformarlos en requisitos técnicos obligatorios, como el protocolo 'Human-in-the-loop'. Esto demostró que el ARA/DPIA es una herramienta de diseño preventivo y no solo de cumplimiento.
3. Valor del Gobierno de Datos (Fase 4): La elaboración del Data Sheet reveló un sesgo crítico en la representatividad de los datos de entrenamiento (80% de un solo proveedor). Sin esta fase, el modelo habría sido entrenado con datos sesgados, llevando a un fallo sistémico en producción. Esto valida la necesidad de una fase dedicada exclusivamente a la gobernanza de los datos antes del desarrollo.

4. Necesidad de Monitoreo Continuo (Fase 8): La operación simulada detectó un 'concept drift' (cambio en el formato de una factura) que no fue anticipado. El sistema de monitoreo y el protocolo de derivación a humanos funcionaron como un control efectivo, demostrando que la gobernanza no termina en el despliegue, sino que es un proceso continuo de vigilancia y adaptación.
5. Importancia de la Supervisión Humana Competente: El incidente INC-2026-003, donde un supervisor humano cometió un error, subrayó que no basta con tener un protocolo de 'Human-in-the-loop'. Es crucial reforzar la capacitación continua de los supervisores para que comprendan las limitaciones del modelo y se puedan aplicar las mejoras que correspondan.

4.3.3.2. Ajustes Finales al Framework:

Basado en estos hallazgos, se realizaron los siguientes ajustes al framework propuesto:

1. Refuerzo del Voto Vinculante del DPO: Se ratificó y se dio mayor énfasis en la documentación a la potestad del DPO para vetar proyectos en el Gate 3 (ARA/DPIA) si las mitigaciones de privacidad no son satisfactorias. La simulación demostró que este es el punto de control más efectivo para proteger los datos personales.
2. Inclusión de Métricas de Calidad de Supervisión Humana: Se añadió al conjunto de KPIs de la Fase 8 un indicador para medir la "Tasa de Corrección de Decisiones de IA por Supervisores", para evaluar la efectividad de la supervisión y detectar necesidades de re-capacitación.
3. Formalización del Despliegue Gradual: La estrategia de despliegue gradual (piloto controlado) demostró ser tan efectiva para mitigar riesgos operativos que se formalizó como una práctica recomendada obligatoria para todos los sistemas de Alto Riesgo dentro de la guía de la Fase 7.
4. Énfasis en el Ciclo de Vida Completo: La simulación de la Fase 9 (Retiro) por obsolescencia tecnológica validó la necesidad de considerar el ciclo de vida completo desde el inicio, incluyendo la planificación para el desmantelamiento seguro de los sistemas y la gestión de datos residuales.

En conclusión, la simulación no solo validó la coherencia y aplicabilidad de la arquitectura propuesta, sino que también permitió refinarla con base en evidencia práctica, resultando en un framework más robusto, pragmático y adaptado a los desafíos reales del sector público.

4.4.Síntesis del Capítulo

El presente capítulo constituye el núcleo operativo de la investigación, transformando los principios normativos en una estructura funcional de gobernanza para el Distrito Capital. A través del paradigma *Design Science Research* (DSR), se ha consolidado un artefacto integral compuesto por una arquitectura de cinco capas y una caja de herramientas operativa, cuya efectividad fue sometida a una validación empírica rigurosa mediante la simulación del **Sistema de Certificados de Residencia**.

Los aspectos más significativos derivados del desarrollo y validación de la propuesta se sintetizan en los siguientes puntos clave:

Integración Arquitectónica y Normativa: La arquitectura de cinco capas (Principios, Gobierno, Ciclo de Vida, Controles y Métricas) logró traducir exitosamente lineamientos abstractos del CONPES 4144 (2024) y el AI Act de la Unión Europea en hitos operativos gestionables. La simulación demostró que la gobernanza no debe ser un apéndice del proyecto, sino el eje que guía el ciclo de vida desde el *Intake* hasta el retiro del sistema.

El Toolkit como Puente de Capacidades: Los instrumentos diseñados (AI Use-Case Canvas, ARA/DPIA, Model Cards, entre otros) probaron ser herramientas esenciales para reducir la brecha técnica en las entidades. Por ejemplo, el uso del **AI Use-Case Canvas** facilitó una alineación estratégica temprana, mientras que el **Checklist de Proveedores** permitió mitigar los riesgos derivados de la dependencia tecnológica externa, estandarizando los requisitos contractuales y de debida diligencia.

Hallazgos Críticos de la Validación Experimental: La simulación del caso de "Alto Riesgo" permitió identificar riesgos que suelen ser invisibles en etapas de diseño teórico. Se destacan tres hallazgos fundamentales:

Detección de Sesgos en el Origen: La aplicación del *Data Sheet* reveló un sesgo crítico de representatividad (80% de datos de un solo proveedor), permitiendo una corrección proactiva antes del entrenamiento.

Mitigación de Sesgos Socioeconómicos: El ARA/DPIA detectó que el motor de OCR fallaba ante imágenes de baja calidad (propias de dispositivos de gama baja), lo que llevó a la implementación obligatoria de un protocolo **Human-in-the-loop** para garantizar el debido proceso y evitar la exclusión de ciudadanos vulnerables.

Identificación de Concept Drift: La fase de monitoreo detectó desviaciones por cambios en los formatos de facturas de servicios públicos, confirmando que la gobernanza es un proceso de vigilancia continua y no un evento de cumplimiento único.

Fortalecimiento de Roles y Controles: La validación ratificó la importancia del voto vinculante del DPO en el Gate 3 como el control más efectivo para la protección de datos personales. Asimismo, se evidenció que la supervisión humana no es solo una formalidad, sino una competencia que requiere capacitación específica para evitar errores de interpretación por parte de los funcionarios.

En síntesis, este capítulo no solo valida la viabilidad técnica y operativa del framework propuesto, sino que demuestra su capacidad para reducir la incertidumbre institucional en Bogotá. Los resultados obtenidos mediante la simulación proporcionan la evidencia empírica necesaria para afirmar que el modelo es aplicable, escalable y capaz de proteger los derechos fundamentales frente a la adopción de tecnologías disruptivas, sentando así las bases para las conclusiones generales del trabajo.

5. Conclusiones

5.1.Conclusiones

El desarrollo de la investigación permitió comprender que Bogotá se encuentra ante un momento decisivo en la adopción de inteligencia artificial dentro del sector público, en función de la disponibilidad de modelos avanzados, el incremento de soluciones basadas en datos y el surgimiento de nuevos marcos regulatorios han generado la necesidad de contar con mecanismos estructurados para el control institucional. La brecha entre el avance tecnológico y la capacidad gubernamental para regular y monitorear su uso se hizo evidente durante el diagnóstico realizado en la Fase 1, que constituyó la base del objetivo específico OE1 (Realizar un diagnóstico integral...), donde se identificaron vacíos metodológicos, falta de lineamientos operativos, dependencia de proveedores externos, ausencia de criterios homogéneos para clasificación de riesgo y debilidad en la trazabilidad del ciclo de vida. La ausencia de herramientas verificables en las entidades distritales representa un desafío que puede afectar derechos ciudadanos, seguridad informática y transparencia administrativa. Esta situación dio origen al objetivo principal de la investigación, cuyo propósito consistió en diseñar, desarrollar y validar un framework de gobernanza que respondiera a dichas necesidades. El resultado confirma la pertinencia del enfoque y su aporte para reducir incertidumbre institucional.

El diseño del framework demostró que es posible articular elementos normativos nacionales, como los definidos en el CONPES 4144 (2024), con estándares internacionales ampliamente reconocidos, tales como el AI Act europeo y el modelo NIST AI RMF, integrando estos referentes con las realidades estructurales del Distrito Capital. La revisión documental permitió extraer principios comunes de gobernanza, seguridad, transparencia, supervisión, documentación técnica, protección de datos y análisis de impacto. Sin embargo, el valor de la investigación no radicó solo en la comparación conceptual, sino en la adaptación contextual y operativa de dichos principios al entorno territorial de Bogotá, lo cual materializa directamente el objetivo específico OE2 (Diseñar la arquitectura integral...). Ese proceso dio origen a una estructura funcional aplicable a entidades con niveles de madurez técnica distintos, recursos limitados y procesos administrativos heterogéneos. Este enfoque territorial demuestra que la gobernanza algorítmica debe construir capacidad institucional a partir de la realidad y no solo desde la referencia normativa.

La arquitectura diseñada en cinco capas se consolidó como un factor estructural decisivo para el éxito del modelo. Cada capa cumple una función específica dentro del enfoque de gobernanza: principios, gobierno, ciclo de vida, controles y métricas. Esta distribución entrega claridad metodológica y facilita el tránsito desde la intención regulatoria hacia la acción operativa. La capa de principios establece el

fundamento ético y jurídico para la toma de decisiones; la capa de gobierno define roles, responsabilidades y mecanismos de aprobación; la capa de ciclo de vida introduce estructura temporal para gestionar etapas desde diseño hasta retiro; la capa de controles ofrece herramientas para medir riesgos, seguridad, sesgo, calidad de datos y trazabilidad; y la capa de métricas permite evaluar impacto institucional y desempeño técnico. El análisis realizado muestra que esta arquitectura fortalece competencias institucionales, reduce discrepancias en decisiones administrativas y aporta coherencia a procesos de adopción tecnológica, consolidando así el logro del OE2.

El valor operativo del framework aumenta significativamente gracias a la caja de herramientas diseñada. Se comprobó que los instrumentos desarrollados permiten aplicar el marco conceptual sin necesidad de capacidades técnicas avanzadas, cumpliendo de esta forma con el objetivo específico OE3 (Desarrollar la caja de herramientas operativa...). Entre las herramientas destacan el AI Use-Case Canvas, que organiza criterios estratégicos para priorizar casos de uso; la matriz de riesgos que permite clasificar impacto institucional bajo criterios técnicos, de seguridad y derechos ciudadanos; el modelo de evaluación ARA o DPIA, que documenta requisitos legales y salvaguardas; el Model Card, que ofrece trazabilidad técnica y descripciones estructuradas; y el checklist de proveedores, que fortalece el control durante procesos de contratación. Estas herramientas fueron diseñadas para responder a uno de los mayores hallazgos del diagnóstico: la necesidad de traducir exigencias normativas en procesos sencillos, auditables y utilizables por equipos multidisciplinares en entidades públicas del Distrito. La estructura de clasificación de riesgos aporta criterios para determinar usos permitidos, condicionados o prohibidos, reforzando seguridad jurídica y prevención de daños.

La investigación también identificó que uno de los desafíos más significativos para Bogotá está relacionado con la dependencia tecnológica de proveedores externos. Esta dependencia es riesgosa cuando no existen criterios de debida diligencia, supervisión técnica del proveedor, documentación de modelos, registro de actualizaciones o mecanismos de verificación contractual. El framework incorpora soluciones para este problema mediante una política de compras y proveedores orientada a estandarizar requisitos contractuales, que forma parte integral de la arquitectura desarrollada bajo el OE2. Este aporte fortalece la integración entre derecho administrativo, seguridad informática y protección de datos, incorporando elementos como prohibiciones de almacenamiento indebido, restricciones de entrenabilidad de modelos, condiciones de confidencialidad, ubicación adecuada de datos y documentación completa. El documento señala que la gestión contractual puede convertirse en una herramienta clave para preservar soberanía tecnológica y evitar riesgos estratégicos.

La validación prevista dentro del marco metodológico demostró que el modelo es viable y aplicable. La Fase 3 de validación experimental, que respondió al objetivo específico OE4 (Validar la efectividad

y usabilidad...), ofreció lineamientos para análisis comparativo, revisión experta, pilotos distritales simulados y ajustes documentales, lo que confirma que el framework no es solo una propuesta conceptual sino un instrumento con vocación de implementación real. La existencia de fases para retroalimentación, revisión metodológica, reescritura, análisis experimental e incorporación de hallazgos emergentes refuerza el valor técnico del modelo. La matriz de roles y responsabilidades del capítulo cuatro expone un proceso organizado para documentar cada etapa y garantizar precisión metodológica, manteniendo coherencia con los objetivos específicos planteados desde el inicio de la investigación.

La evaluación institucional realizada en el diagnóstico confirmó que el Distrito Capital necesita estrategias para fortalecer transparencia, trazabilidad y control de decisiones automatizadas. El framework responde a esta necesidad al integrar mecanismos explícitos de documentación técnica, supervisión humana, auditoría de uso, ciclo de aprobación, clasificación de riesgos, gestión de incidentes y monitoreo permanente, elementos que fueron diseñados y consensuados durante el cumplimiento del OE2 y OE3. Estas recomendaciones se alinean con la matriz del documento, donde se establece que casos de uso con impacto jurídico, decisiones automáticas o datos sensibles deben considerarse de alto riesgo e incluso prohibirse en determinadas condiciones. El modelo propone restricciones claras que reducen margen de error institucional, fortalecen confianza pública y protegen derechos fundamentales.

El análisis permitió observar que la gobernanza algorítmica no depende únicamente de elementos jurídicos o técnicos aislados, sino de la interacción entre ambos. La ausencia de criterios institucionales puede conducir a decisiones administrativas poco sólidas, modelos opacos, fallas de documentación, sesgos no controlados y riesgos significativos para integridad institucional. La investigación demostró que el framework desarrollado permite reducir esa vulnerabilidad mediante procesos estandarizados que incorporan criterios jurídicos de protección de datos, principios de equidad, lineamientos de transparencia, funciones de auditoría y métricas de desempeño. Esta integración aporta orden, estructura y consistencia a decisiones técnicas complejas, evidenciando la consecución del OE2 y OE3.

El trabajo confirmó que la adopción de IA en el sector público requiere herramientas para garantizar que la supervisión humana mantenga el control sobre decisiones automatizadas. El framework establece procedimientos para asegurar intervención humana en funciones clave cuando existe impacto sobre derechos ciudadanos, reputación institucional o decisiones administrativas sensibles. La introducción de roles de supervisión, revisión posterior, ciclos de aprobación y monitoreo contribuye a evitar automatización ciega, una condición que se ha identificado como riesgo en distintos

marcos internacionales y que fue abordada específicamente en el diseño de la Capa 4 del framework (OE2).

Otro resultado importante de la investigación fue la identificación de mecanismos para gestionar incidentes asociados al uso de sistemas de IA. El documento señala que, en caso de materializarse un riesgo, las entidades deben cerrar el flujo de procesamiento, notificar al delegado de protección de datos, registrar el hecho, documentar el incidente y activar procesos de revisión institucional. Este lineamiento fortalece capacidad de respuesta pública y crea condiciones para corregir fallas antes de que generen daños estructurales. La incorporación de este procedimiento, detallado en el toolkit desarrollado (OE3), demuestra que la gobernanza de IA no se limita a prevención, sino también a mitigación y aprendizaje.

A nivel estratégico, el framework posiciona al Distrito Capital como un actor capaz de liderar procesos de gobernanza tecnológica en el ámbito territorial. La investigación demuestra que Bogotá cuenta con condiciones institucionales, profesionales y normativas para convertirse en referente para gobiernos locales, regionales y nacionales, al ofrecer un modelo orientado a derechos, sostenido por evidencia y compatible con prácticas reconocidas internacionalmente. El diseño del framework no se limita a resolver una necesidad local inmediata, sino que representa un avance conceptual con potencial de expansión para fortalecer gobernanza en diferentes sectores del Estado, lo cual fue consolidado mediante la Fase 4 y el cumplimiento del OE5 (Consolidar los hallazgos y elaborar la guía de implementación...).

El proceso permitió consolidar un instrumento robusto que responde directamente a los objetivos planteados. El diagnóstico (OE1) identificó brechas estructurales; la sistematización organizó capacidades, requisitos y hallazgos; la selección comparativa permitió validar componentes de marcos internacionales; el diseño del framework (OE2) estableció estructura funcional y operativa; el desarrollo del toolkit (OE3) tradujo el modelo en artefactos de trabajo; y la validación (OE4) consolidó pertinencia institucional y viabilidad técnica. La fase final de consolidación (OE5) integró todos estos elementos en una propuesta completa y transferible. Este proceso demuestra el cumplimiento pleno del objetivo principal [OP] y de los cinco objetivos específicos establecidos en el capítulo tercero y confirma que el modelo construido no es teórico sino técnico, operativo y aplicable.

5.2. Líneas de Trabajo Futuro

El desarrollo del framework de gobernanza para sistemas de inteligencia artificial aplicado al contexto distrital abre un panorama amplio de investigación, implementación, expansión institucional,

fortalecimiento técnico y evolución conceptual. El modelo diseñado demostró viabilidad operativa, amplitud estructural y capacidad de adaptación, y se convierte hoy en un punto de partida para múltiples agendas futuras. Los resultados alcanzados no representan un cierre temático, sino un inicio que habilita procesos sostenibles de transformación pública. De este modo, las líneas de trabajo futuro constituyen una oportunidad para perfeccionar el modelo, extender su uso a nuevos entornos, elevar su impacto social, fortalecer validación técnica, incorporar avances normativos emergentes y construir un sistema progresivo de gobernanza algorítmica en Colombia.

El avance acelerado de la inteligencia artificial ha modificado las dinámicas tecnológicas, jurídicas, económicas y sociales del mundo público. Esto implica que ningún framework es definitivo. La gobernanza algorítmica es evolutiva por naturaleza, pues responde a cambios constantes en metodologías, arquitecturas de modelos, estructuras de datos, regulaciones internacionales, expectativas sociales y capacidades institucionales. Por lo tanto, el modelo propuesto debe ser entendido como un instrumento vivo que crecerá con el contexto, con las tecnologías y con las necesidades del país.

5.2.1. Aplicación territorial y fortalecimiento distrital

Una primera dirección de crecimiento consiste en extender el framework a más entidades distritales, no solo para repetir resultados, sino para generar un mapa de variaciones operativas. El despliegue permitirá examinar diferencias entre sectores administrativos, capacidades técnicas, estructuras organizacionales y niveles de riesgo. En salud, por ejemplo, el marco podría guiar el análisis de modelos predictivos orientados a vigilancia epidemiológica, triage digital o priorización de servicios públicos. En movilidad permitiría evaluar sistemas algorítmicos para optimizar rutas, clasificar eventos viales, priorizar intervenciones o analizar patrones de congestión. En hacienda podría apoyar la detección automatizada de evasión tributaria, la predicción financiera o la clasificación de transacciones.

En cada escenario surgirán nuevos aprendizajes. Algunas entidades requerirán fortalecer procesos de documentación, otras necesitarán medidas adicionales para garantizar explicabilidad, mientras que otras deberán adoptar métricas más avanzadas para evaluar desempeño, eficiencia y equidad. Este proceso nutrirá el marco de retroalimentación real, perfeccionará la matriz de riesgos y validará fortalezas y oportunidades de mejora.

El trabajo futuro también podrá centrarse en el diseño de indicadores de madurez distrital. Un índice comparativo permitiría calificar niveles de avance por entidad, determinar brechas estructurales, orientar asignación de recursos y ofrecer un lenguaje común para toma de decisiones. Este índice podrá ser utilizado de manera periódica para monitorear evolución, garantizar mejoras continuas y prevenir retrocesos.

5.2.2. Expansión a niveles de gobierno nacionales y locales fuera de Bogotá

Una línea futura clave consiste en escalar el modelo a gobiernos locales y entidades nacionales. El framework ofrece estructura central para construir un sistema de gobernanza algorítmica fuera del territorio distrital, manteniendo principios y controles esenciales. Esta evolución permitiría recoger resultados desde ciudades, departamentos, agencias estatales y ministerios.

Un proyecto de adopción nacional requerirá análisis comparativo entre condiciones territoriales diversas, niveles de digitalización diferentes, y capacidades institucionales heterogéneas. Ello implicará diseñar herramientas de adaptación sectorial, flexibilizar requisitos del toolkit, incorporar recomendaciones de política pública y analizar modelos de descentralización técnica.

La expansión territorial facilitaría avanzar hacia un ecosistema colombiano de gobernanza basada en IA. Esta construcción interinstitucional permitiría: crear marcos compartidos de responsabilidad administrativa, estructurar inventarios nacionales de algoritmos, consolidar métricas de desempeño, generar estándares unificados de transparencia y fortalecer control público sobre sistemas automatizados. Un marco de esa naturaleza aportaría a la transformación digital del Estado colombiano y aumentaría el grado de confianza ciudadana frente al uso de IA gubernamental.

5.2.3. Apertura sectorial hacia el entorno privado

Otra línea relevante será adaptar el framework para su uso en empresas privadas. La adopción del modelo permitiría fortalecer trazabilidad algorítmica en organizaciones con uso intensivo de datos, automatización y decisiones automatizadas. Industrias como banca, seguros, transporte, educación, salud privada y logística podrían beneficiarse mediante aplicación de criterios técnicos para gestión de riesgo, protección de datos, transparencia algorítmica y equidad.

Este esfuerzo generará oportunidades de investigación aplicada en temas como responsabilidad empresarial, auditoría técnica de modelos, justicia algorítmica, interacción con usuarios finales, calidad de datos, gestión de sesgos, sistemas de reputación tecnológica y transparencia contractual. El desarrollo de esta línea también permitiría promover colaboraciones público privadas orientadas a eficiencia operativa y mejora de servicios con impacto directo en ciudadanía.

5.2.4. Desarrollo metodológico basado en pilotos y análisis comparativo

Una dirección futura de gran potencial está relacionada con la experimentación controlada en entidades reales. La validación del framework podrá incluir pilotos temáticos para medir eficiencia, trazabilidad y desempeño técnico. Cada piloto representará un caso de estudio orientado a documentar resultados, comparar variaciones, identificar limitaciones, medir riesgos y producir recomendaciones.

Los resultados de estos pilotos permitirán mejorar matriz de riesgos, enriquecer el toolkit, refinar flujos de aprobación, fortalecer procedimientos de auditoría y ajustar requerimientos de supervisión humana. El análisis comparativo entre pilotos facilitará establecer patrones de riesgo, identificar tendencia de incidentes y documentar metodologías de mejora para entidades públicas que recién inician procesos de adopción.

5.2.5. Aplicación en el contexto legal, regulatorio y normativo

Una línea estratégica fundamental será profundizar en las implicaciones jurídicas del framework. El modelo tiene capacidad para convertirse en referencia legal dentro del país, incorporando elementos que puedan integrarse a futuras políticas públicas, lineamientos regulatorios, decretos administrativos o guías técnicas obligatorias. La investigación abre espacio para construir argumentación comparativa con marcos internacionales, analizar evolución normativa nacional, interpretar tensiones jurídicas derivadas del uso de IA y diseñar recomendaciones para legisladores, jueces, entidades de control y organismos reguladores.

El área normativa podrá incluir el análisis detallado del AI Act europeo, la exploración de categorías de riesgo equivalentes, la adaptación de procesos de certificación algorítmica, la definición de restricciones para sistemas de alto impacto, la propuesta de requisitos de transparencia obligatoria y la incorporación de directrices vinculantes en compras públicas. También será posible desarrollar protocolos reglamentarios para gestión de incidentes algorítmicos, definir derechos ciudadanos ante sistemas de IA y delimitar responsabilidad administrativa frente a fallas automatizadas.

El sector legal podrá utilizar el framework como guía interpretativa para resolver futuros debates jurídicos, como la validez probatoria de decisiones automatizadas, la responsabilidad por errores derivados de modelos predictivos, la protección de datos personales en algoritmos, las garantías del debido proceso en sistemas automatizados y el impacto de la supervisión algorítmica sobre derechos fundamentales.

Esta línea de crecimiento permitirá posicionar el modelo como instrumento jurídico de referencia, fortaleciendo seguridad institucional, confianza ciudadana y coherencia normativa frente a un campo tecnológico en evolución permanente.

REFERENCIAS BIBLIOGRÁFICAS

Access Now. (2023). *Regulatory mapping on artificial intelligence in Latin America*. Access Now.

Ada Lovelace Institute. (2022). *Algorithmic impact assessment: A case study in healthcare*. Ada Lovelace Institute.

Alpaslan, G. (2025). Exploring the impact of ISO/IEC 42001:2023 AI management system. *Anadolu Akademi International Refereed Journal of Social Sciences (AAIRJ)*, 7(2), 1-14.

ALSUR. (2024). *The regulatory pathways for AI in Latin America*. ALSUR.

Castaño Castaño, S. (2025). La inteligencia artificial en Salud Pública: oportunidades, retos éticos y perspectivas futuras. *Revista Española de Salud Pública*, *99*(1), e202503017. https://ojs.sanidad.gob.es/index.php/resp/article/view/1006

CLAD. (2024). Inteligencia Artificial en el Sector Público latinoamericano. *Revista del CLAD Reforma y Democracia*, *88*, 116--143.

Comisión Europea. (2025). *AI Act | Shaping Europe's digital future*. European Commission.

Cotino, L., y Castellanos, J. (Eds.). (2023). *Algoritmos abiertos y que no discriminen en el sector público*. Tirant lo Blanch.

Congreso de la República de Colombia. (2022, 21 de diciembre). *Ley 2279 de 2022. Por la cual se decreta el Presupuesto del Sistema General de Regalías para el bienio del 1o. de enero de 2023 al 31 de diciembre de 2024*. Diario Oficial No. 52.255.

Consejo Nacional de Política Económica y Social. (2019). *Documento CONPES 3975: Política nacional para la transformación digital e inteligencia artificial*. Departamento Nacional de Planeación. https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3975.pdf

Consejo Nacional de Política Económica y Social. (2024). *Documento CONPES 4144: Hoja de ruta de la inteligencia artificial para Colombia: retos actuales para una transformación futura*. Departamento Nacional de Planeación. https://www.dnp.gov.co/publicaciones/Planeacion/Paginas/conpes-4144-hoja-de-ruta-colombia-inteligencia-artificial-retos-actuales-transformacion-futura.aspx

Departamento Nacional de Planeación. (2022). *Manual del Plan Estratégico de Tecnologías de la Información – PETI 2022-2026*. https://www.dnp.gov.co/

Departamento Nacional de Planeación. (2025). *CONPES 4144: Hoja de ruta Colombia inteligencia artificial - retos actuales y transformación futura*. https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/4144.pdf

European Commission. (2021, April 21*). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* COM(2021) 206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Evans, P. (1995). *Embedded autonomy: States and industrial transformation*. Princeton University Press.

Fernández Torres, R. (2025). Impacto de la inteligencia artificial en la gestión tributaria de las PYMES: avances, desafíos y oportunidades en México. *Revista Latinoamericana de Tecnología*, *4*(2), 123-145. https://latam.redilat.org/index.php/lt/article/view/3761

Filgueiras, F. (2023a). Desafíos de gobernanza de inteligencia artificial en América Latina. Infraestructura, descolonización y nueva dependencia. *Revista del CLAD Reforma y Democracia*, (87), 5-36. https://doi.org/10.69733/clad.ryd.n87.a3

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

G7 Digital and Technology Ministers. (2023). *G7 Hiroshima AI Process: International Guiding Principles and Code of Conduct for AI*. https://digital-strategy.ec.europa.eu/en/news/commission-welcomes-g7-leaders-agreement-guiding-principles-and-code-conduct-artificial

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for Datasets. *Communications of the ACM*, 64(12), 86–92. https://doi.org/10.1145/3458723

González Méndez, A., y Vásquez López, P. (2024). Trayectoria y modelo de gobernanza de las políticas de inteligencia artificial de los países de América del Norte. *Revista Justicia*, *29*(45), 78-102. https://revistas.unisimon.edu.co/index.php/justicia/article/view/7162

González Fuster, G. (2020). *Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights*. European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs.

Gutiérrez Peña, M. (2024). Balance de la política pública de inteligencia artificial y transformación digital (2019-2024). *Revista de Investigación de Pensamiento Liberal*, *2*(3), 45-68. https://revistapensamientoliberal.com/index.php/RIPL/article/view/44

Heeks, R. (2002). Information systems and developing countries: Failure, success, and local improvisations. *The Information Society*, *18*(2), 101-112. https://doi.org/10.1080/01972240290075039

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Instituto Colombiano de Normas Técnicas y Certificación. (2023). *Marco ético para la inteligencia artificial en Colombia*. Presidencia de la República de Colombia.

International Organization for Standardization. (2023a). *ISO/IEC 23894:2023 - Artificial Intelligence - Risk management*. https://www.iso.org/standard/77304.html

International Organization for Standardization. (2023b). *ISO/IEC 42001:2023 - Information technology - Artificial intelligence - Management system*. https://www.iso.org/standard/42001

Internet Policy Review. (2024). The impact of AI automation on social benefit provision in Brazil. *Internet Policy Review*, *13*(3), 1--20.

ISACA. (2018). COBIT 2019 Framework: Governance and Management Objectives. Information Systems Audit and Control Association.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report EBSE-2007-01). Keele University and Durham University.

Congreso de la República de Colombia. (2012, 17 de octubre). Ley 1581 de 2012. Por la cual se dictan disposiciones generales para la protección de datos personales. Diario Oficial No. 48.587. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981

López, M. (2021). Inteligencia artificial (IA) aplicada a la gestión pública. *Revista Venezolana de Gerencia*, *26*(94), 1100-1118. https://produccioncientificaluz.org/index.php/rvg/article/view/35767

Martínez González, J., Rodríguez Pérez, L., y Silva Hernández, M. (2024). Percepción de la inteligencia artificial en la lucha contra la corrupción: una exploración al caso del Estado de Colombia. *OPERA*, *35*, 187-214. https://revistas.uexternado.edu.co/index.php/opera/article/view/9971

Mendonça, R. F., Filgueiras, F., & Almeida, V. A. (2023). *Algorithmic institutionalism: The changing rules of social and political life*. Oxford University Press. https://doi.org/10.1093/oso/9780192870070.001.0001

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19) (pp. 220–229). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287596

National Institute of Standards and Technology. (2023). *AI Risk Management Framework (AI RMF 1.0)*. https://www.nist.gov/itl/ai-risk-management-framework

National Institute of Standards and Technology. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence

OECD. (2022). *OECD Framework for the Classification of AI Systems*. OECD Digital Economy Papers, No. 323. OECD Publishing. https://doi.org/10.1787/cb615418-en

OECD. (2024). *Governing with Artificial Intelligence: Implementation challenges that hinder the strategic use of AI in government*. OECD Publishing.

OECD & IDB. (2024). *2023 OECD/IDB Digital Government Index of Latin America and the Caribbean*. OECD Publishing.

OECD OPSI. (2023). *Chile's road to algorithmic transparency: Setting new standards*. Observatory of Public Sector Innovation.

OEA. (2025). *Marco Interamericano de Gobernanza de Datos e Inteligencia Artificial (MIGDIA)*. Organización de los Estados Americanos.

OGP Colombia. (2025). *Crear un modelo de gobernanza para la inteligencia artificial en Colombia*. Open Government Partnership.

Organisation for Economic Co-operation and Development. (2024). *OECD AI Principles*. https://oecd.ai/en/ai-principles

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.

Presidencia de la República de Colombia. (2013, 27 de junio). Decreto 1377 de 2013. Por el cual se reglamenta parcialmente la Ley 1581 de 2012. Diario Oficial No. 48.834. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=53646

Restrepo Silva, A., y González Vargas, P. (2024). *Perfil de las asociaciones público-privadas en servicios e infraestructura de salud de América Latina y el Caribe: Principales cifras y tendencias del sector*. Banco Interamericano de Desarrollo.

Revista CLAD. (2024). Inteligencia artificial en el sector público latinoamericano. Estudio comparado a partir de la Carta Iberoamericana de Inteligencia Artificial en la Administración Pública. *Revista del CLAD Reforma y Democracia*, *89*, 123-156. https://revista.clad.org/ryd/article/view/387

Rivera Hernández, S. (2024). Inteligencia artificial en el sector público en México. *New Business Review*, *2*(4), 89-112. https://journals.epnewman.edu.pe/index.php/NBR/article/view/371

Ruiz Sánchez, D. (2024). Implementación de la Inteligencia Artificial en las Altas Cortes de Colombia: los casos de la Corte Constitucional y el Consejo de Estado. *REDOEDA*, *12*(1), 156-189. https://bibliotecavirtual.unl.edu.ar/publicaciones/index.php/Redoeda/article/view/13824

Superintendencia de Industria y Comercio [SIC]. (2024, 21 de agosto). *Circular Externa No. 002 de 2024. Lineamientos sobre el Tratamiento de Datos Personales en Sistemas de Inteligencia Artificial*. https://sedeelectronica.sic.gov.co/

Uñate, L., y Hernández, A. (2022). Nudging e inteligencia artificial contra la corrupción en el sector público: posibilidades y riesgos. *Derecho del Estado*, *52*, 187-214. https://revistas.uexternado.edu.co/index.php/Deradm/article/view/7859

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

UNESCO. (2023). *Guidance for generative AI in education and research*. https://unesdoc.unesco.org/ark:/48223/pf0000386693

UNESCO. (2025). *UNESCO's global AI training empowers civil servants to revolutionize public services*. UNESCO.

Vestri, G. (2024). La fusión transformadora entre el sector público y la inteligencia artificial: el "test de evaluación de impacto" como prioridad. *International Journal of Digital Law*, 4(3), 43–64. https://doi.org/10.47975/digital.law.vol.4.n.3.vestri

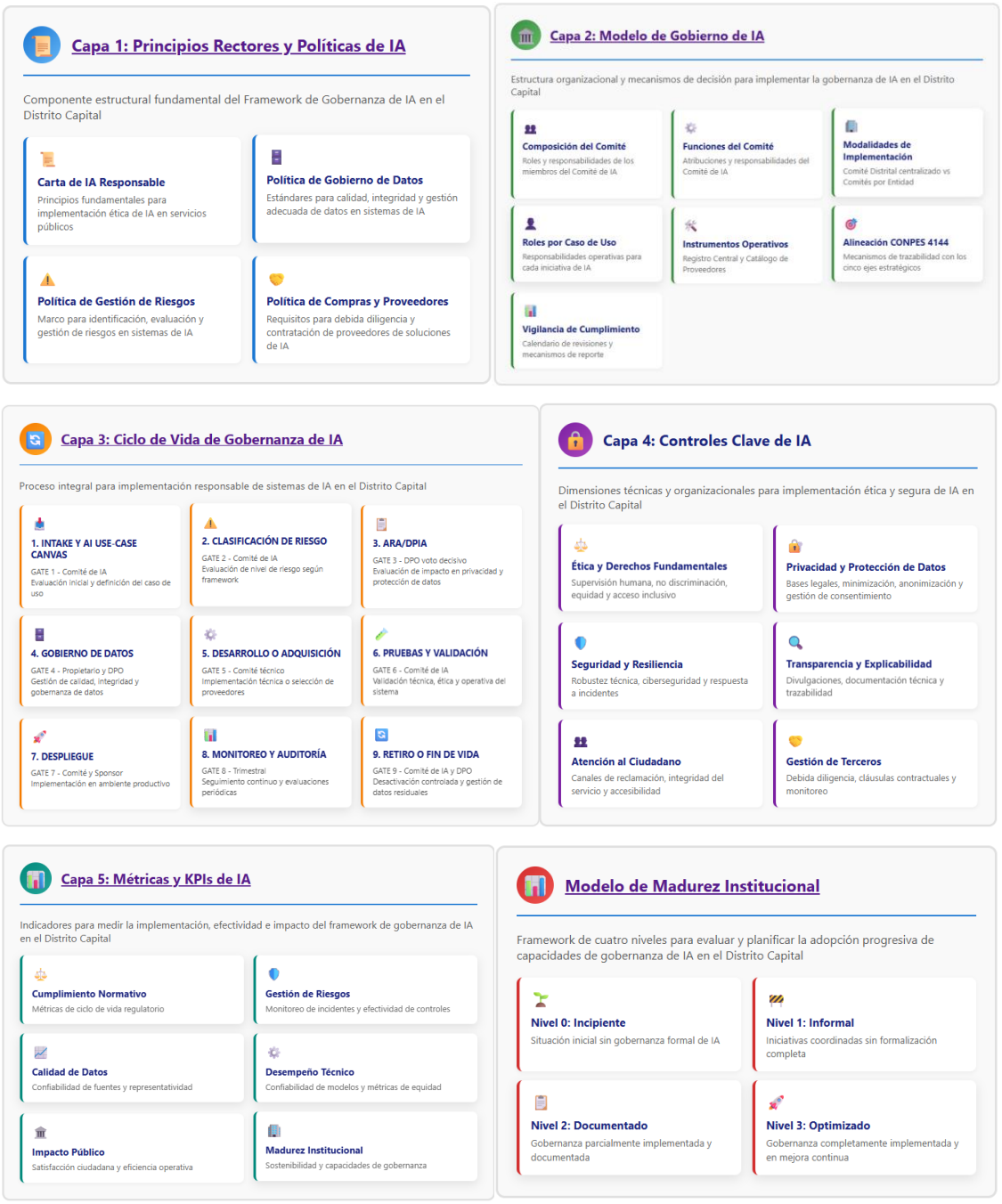
Zamora Pérez, A. L. (2025). *Inteligencia Artificial en la Administración Pública, sus Retos, Oportunidades y Casos en América Latina y Ecuador*. Revista Internacional de Investigación y Desarrollo Global, 4(3), 47-60. https://doi.org/10.64041/riidg.v4i3.48

ANEXO A. TABLA COMPARATIVA AVANZADA DE FRAMEWORKS

Framework	**Cobertura normativa**	**Adaptabilidad contextual**	**Herramientas prácticas**	**Métricas de implementación**	**Limitaciones subnacionales**
AI Act (UE)	Regulación integral orientada a riesgo con obligaciones, documentación y supervisión para sistemas de alto riesgo y GPAI (Comisión Europea, 2025)	Depende de autoridades competentes y coordinación multinivel para operatividad territorial (Comisión Europea, 2025)	Evaluaciones de conformidad, documentación técnica y registros centralizados (Comisión Europea, 2025)	Plazos y verificación obligatoria con capacidad sancionatoria (Comisión Europea, 2025)	Altas exigencias técnicas y de certificación frente a recursos locales (Comisión Europea, 2025)
NIST AI RMF (2023)	Marco voluntario con funciones Govern, Map, Measure y Manage y perfiles adaptables (NIST, 2023)	Flexible y portable por caso de uso con guías por perfiles y playbooks (NIST, 2023)	Catálogos de resultados, prácticas sugeridas y materiales de apoyo (NIST, 2023)	Métricas sugeridas no obligatorias y énfasis en mejora continua (NIST, 2023)	Presupone experticia y cultura de gestión de riesgo especializada (NIST, 2023)
UNESCO Recomendación	Estándar normativo en ética de IA con enfoque de derechos humanos (UNESCO, 2024)	Apuesta por adaptación cultural e institucional y formación del servicio civil (UNESCO, 2024)	Orientaciones generales y programas de capacitación para lo público (UNESCO, 2024)	Sin indicadores técnicos detallados en el propio instrumento (UNESCO, 2024)	Riesgo de quedarse en lineamientos generales sin guías técnicas finas (UNESCO, 2024)
Principios OCDE	Principios y herramientas para políticas de IA y gobierno digital con observatorios comparados (OECD, 2024)	Alta flexibilidad para adaptación nacional y subnacional (OECD, 2024)	Guías, casos y diagnósticos regionales para capacidades públicas (OECD, 2024)	Indicadores de madurez y brechas sistémicas (OECD, 2024)	Enfoque comparado general con referencias de alta capacidad (OECD, 2024)
ISO/IEC 42001 (2023)	Sistema de gestión certificable con controles, auditorías y revisión de desempeño (Alpaslan, 2025)	Aplicabilidad transversal con adaptación por procesos (Alpaslan, 2025)	Procedimientos, evaluaciones internas y auditoría externa (Alpaslan, 2025)	Métricas ligadas a certificación y mejora continua (Alpaslan, 2025)	Costos de certificación y peritaje externo elevados (Alpaslan 2025)

Fuente: Elaboración Propia

ANEXO B. ARQUITECTURA DEL FRAMEWORK DE GOBERNANZA DE IA



Fuente: Elaboración Propia

ANEXO C. DEFINICIÓN DE DASHBOARD PARA MÉTRICAS Y KPIS

DIMENSIÓN	MÉTRICA	KPI ASOCIADO	FRECUENCIA MEDICIÓN	VINCULACIÓN NORMATIVA
CUMPLIMIENTO NORMATIVO	Porcentaje de casos de uso registrados	(Casos de uso registrados en el Registro Central / Total de casos de uso identificados en operación o desarrollo) × 100	Mensual, con auditoría trimestral de completitud	CONPES 4144: Trazabilidad del ecosistema de IA; Ley 2279/2022: Transparencia en tecnologías emergentes
	Cobertura de uso del Catálogo Distrital de Proveedores Pre-evaluados	(Número de contrataciones que utilizaron proveedores del catálogo / Total de contrataciones de sistemas de IA durante el período) × 100	Trimestral, con consolidación anual	Ley 2279/2022: Eficiencia en contratación pública; CONPES 4144: Optimización de procesos
	Porcentaje de casos de uso con clasificación de riesgo documentada	(Casos de uso con clasificación de riesgo formalizada / Total de casos de uso de IA en operación o desarrollo) × 100	Trimestral, con revisión completa anual	CONPES 4144: Gestión basada en riesgos como eje transversal; AI Act UE: Modelo de niveles de riesgo
	Porcentaje de sistemas de alto riesgo con ARA/DPIA completo	(Sistemas de alto riesgo con ARA/DPIA realizado según plantilla del toolkit / Total de sistemas de alto riesgo) × 100	Verificación continua antes de despliegue, consolidado trimestral	Circular Externa 002-2024 SIC: Obligatoriedad ARA para tratamientos de datos personales; CONPES 4144
	Porcentaje de sistemas con documentación técnica completa	(Sistemas con Model Card y Data Sheets actualizados / Total de sistemas en operación) × 100	Semestral	CONPES 4144: Transparencia y auditabilidad; AI Act: Documentación técnica obligatoria
	Cobertura de capacitación obligatoria en IA Responsable y Generativa	(Funcionarios certificados vigentemente en IA Responsable y/o IA Generativa según rol / Total de funcionarios en roles clave + usuarios activos de herramientas GenAI) × 100	Trimestral, con verificación continua integrada a sistemas de autenticación	CONPES 4144: Desarrollo de capacidades como habilitador crítico; Ley 2279/2022: Alfabetización digital en tecnologías emergentes
GESTIÓN DE RIESGOS	Número de incidentes de IA por tipología	Conteo mensual desagregado por tipo: técnico, privacidad, equidad, seguridad, quejas ciudadanas	Registro continuo, reporte consolidado mensual	COBIT 2019 DSS01 (Manage Operations): Monitoreo continuo del desempeño operativo
	Tiempo medio de respuesta a incidentes	Σ(tiempo desde detección hasta resolución completa) / Número total de incidentes procesados	Consolidación mensual con alertas automatizadas	NIST AI RMF: Gestión proactiva de riesgos; COBIT 2019: Cumplimiento de SLA de respuesta
	Tasa de materialización de riesgos	(Riesgos que efectivamente se presentaron durante período / Total de riesgos identificados en ARADPIA durante período anterior) × 100	Semestral	NIST AI RMF: Efectividad de mitigaciones; ISO/IEC 23894: Validación de análisis de riesgos
	Efectividad de controles implementados	% de controles evaluados como efectivos en auditoría	Anual auditoría interna, bianual externa para alto riesgo	COBIT 2019 MEA (Monitor, Evaluate and Assess): Evaluación periódica de controles
CALIDAD DE DATOS	Completitud de datos	(Registros sin valores faltantes en campos críticos / Total de registros en dataset) × 100	Evaluación mensual para datasets en uso activo	CONPES 4144: Infraestructura de datos de

DIMENSIÓN	MÉTRICA	KPI ASOCIADO	FRECUENCIA MEDICIÓN	VINCULACIÓN NORMATIVA
				calidad como pilar fundamental
	Tasa de detección de drift	Número de casos donde se detectó drift significativo durante monitoreo	Semanal alto riesgo, mensual riesgo limitado	ISO/IEC 23894: Monitoreo de degradación de desempeño; CONPES 4144: Mantenimiento de sistemas
	Representatividad demográfica	Test Chi-cuadrado: p-valor ≥ 0.05 sin diferencia significativa	Trimestral o con actualización significativa de datos	AI Act: Equidad en sistemas de IA; CONPES 4144: No discriminación como principio fundamental
	Índice de calidad de documentación de datos	Puntuación 0-100 basada en checklist integral de Data Sheets	Con cada nuevo conjunto de datos o actualización importante	AI Act: Requisitos de documentación; ISO/IEC 42001: Transparencia en gestión de datos
DESEMPEÑO TÉCNICO	Precisión y desempeño de modelos	Clasificación: Precisión, Recall, F1-Score, AUC-ROC; Regresión: MAE, RMSE, R^2; Generación: BLEU, ROUGE, evaluación humana	Monitoreo continuo con alertas automatizadas, reporte semanal/mensual	NIST AI RMF: Medición de desempeño; CONPES 4144: Confiabilidad de sistemas de IA
	Métricas de equidad (fairness)	Disparate Impact Ratio: (Tasa resultado positivo grupo protegido)/(Tasa resultado positivo grupo mayoritario); Equal Opportunity Difference: diferencia absoluta en tasas de acierto	Semanal o mensual en monitoreo continuo	AI Act: Equidad como requisito fundamental; CONPES 4144: No discriminación en sistemas automatizados
	Disponibilidad del sistema	(Tiempo de operación efectiva / Tiempo total planificado) × 100	Monitoreo continuo, reporte consolidado mensual	COBIT 2019 DSS: Continuidad del servicio; Constitución 1991: Derecho a servicios públicos eficientes
	Tiempo de respuesta de sistemas	Percentil 95 del tiempo de respuesta a consultas o inferencias del modelo	Monitoreo continuo, reporte semanal	CONPES 4144: Experiencia ciudadana en gobierno digital; DAFP: Calidad en servicios digitales
IMPACTO EN SERVICIO PÚBLICO	Satisfacción ciudadana	CSAT: % satisfecho/muy satisfecho escala 1-5; NPS: (% promotores - % detractores)	Captura continua mediante muestreo estadístico, consolidación mensual	CONPES 4144: Valor público como objetivo; DAFP: Impacto medible en servicio público
	Eficiencia operativa	(Tiempo post-IA / Tiempo pre-IA - 1) × 100 o (Costo post-IA / Costo pre-IA - 1) × 100	Evaluación trimestral	CONPES 4144: Eficiencia y productividad como objetivos clave de IA en sector público
	Volumen de servicio y tasa de resolución	Número transacciones procesadas; (Casos resueltos sin escalamiento a humano / Total casos procesados) × 100	Consolidación mensual	CONPES 4144: Escalabilidad y eficiencia operativa; DAFP: Modernización administrativa
	Tasa de escalamiento a humano	(Casos escalados a revisión o decisión humana / Total de casos procesados) × 100	Consolidación semanal/mensual	COBIT 2019 DSS01: Balance entre autonomía y

DIMENSIÓN	MÉTRICA	KPI ASOCIADO	FRECUENCIA MEDICIÓN	VINCULACIÓN NORMATIVA
				supervisión; CONPES 4144: Control humano apropiado
	Equidad de acceso	Análisis desagregado de uso por grupos demográficos (género, edad, estrato, localidad)	Evaluación trimestral	Constitución 1991: Igualdad y no discriminación; CONPES 4144: Inclusión digital y equidad de acceso
MADUREZ INSTITUCIONAL	Nivel de madurez institucional	0 (Incipiente), 1 (Informal), 2 (Documentado), 3 (Optimizado)	Evaluación anual con autoevaluación y validación externa	COBIT 2019: Niveles de capacidad 0-5; CONPES 4144: Progresión organizacional en gobernanza de IA
	Cobertura de capacitación en roles clave	(Capacitados roles clave / Total roles clave) × 100	Semestral	CONPES 4144: Talento humano como habilitador estratégico; Ley 2279/2022: Desarrollo de capacidades
	Porcentaje de trámites y servicios con IA implementada	(Servicios con componentes de IA / Total de servicios candidatos identificados en roadmap) × 100	Evaluación semestral	DAFP, MinTIC: Transformación digital progresiva; CONPES 4144: Hoja de ruta de implementación
	Relación costo-beneficio	(Beneficios tangibles monetarios + valor estimado de beneficios intangibles) / (Costos totales de implementación 3 años)	Evaluación anual	DAFP: Sostenibilidad fiscal de inversiones en TI; CONPES 4144: Retorno de inversión en tecnología
	Índice de sostenibilidad institucional	Evaluación 0-100: apropiación, continuidad presupuestal, independencia, transferencia, capacidad de autosuficiencia en capacitación	Anual	COBIT 2019: Sostenibilidad de procesos; DAFP: Continuidad institucional de iniciativas de transformación

Fuente: Elaboración Propia

ANEXO D.GUÍA DE IMPLEMENTACIÓN DEL CICLO DE VIDA DE GOBERNANZA DE IA

Guía de Implementación de Flujo de Gobernanza de IA

Ir a Link: Ciclo de vida IA

Nueve etapas sucesivas para garantizar una adopción responsable y estructurada de sistemas de IA.

Fase 1 - Intake y AI Use-Case Canvas

Descripción: Se formaliza la idea del caso de uso mediante el AI Use-Case Canvas.

Actividades:

- Completar AI Use-Case Canvas.
- Valoración de viabilidad.
- Identificación de riesgos y partes interesadas.

Gate de Revisión G1 - Comité de IA: El Comité de IA evalúa la alineación estratégica, viabilidad y propósito.

Entregable: AI Use-Case Canvas aprobado.

Herramientas: IA Use-Case Canvas

Guía de Implementación: Fase 1 - Intake y AI Use-Case Canvas

Esta fase es el punto de partida formal para cualquier iniciativa de IA. Es liderada por el área de negocio o usuaria que identifica una oportunidad o necesidad.

Matriz RACI para esta fase:

- **Comité de IA:** Consultado (C)
- **DPO:** Consultado (C)
- **Responsable Técnico:** Responsible (R)
- **Sponsor de Negocio:** Accountable (A)
- **Área Jurídica:** Consultado (C)

Actividades Clave y sus Referencias en el Framework

1. Completar AI Use-Case Canvas

Descripción de la Actividad: El Sponsor de Negocio (el líder del área usuaria) es el responsable de liderar la creación de la propuesta. Utilizando la plantilla del AI Use-Case Canvas, debe documentar de manera clara y concisa las 12 secciones que componen la herramienta. Esto incluye:

- El problema a resolver y los objetivos que se persiguen.
- Los actores involucrados (usuarios, ciudadanos afectados).
- Los datos que se necesitarían, haciendo una mención inicial si son personales o sensibles.
- Los riesgos preliminares y las métricas de éxito esperadas.
- Una primera estimación del plan de despliegue y monitoreo.

Referencias clave (Capítulo 4.1 y 4.2):

- **Caja de Herramientas (Sección 4.2.1):** Esta actividad es la aplicación directa del instrumento "AI Use-Case Canvas". El diligenciamiento de la plantilla obliga al proponente a pensar en todas las dimensiones de la gobernanza desde el primer momento.
- **Modelo de Gobierno (Sección 4.1.2.2):** Es la primera gran responsabilidad del rol de "Sponsor de Negocio (Product Owner)", que es quien impulsa la iniciativa.
- **Principios Fundamentales (Sección 4.1.1):** El canvas fuerza a una reflexión temprana sobre todos los principios. Por ejemplo, la sección de "Datos Requeridos" activa el principio de "Privacidad por diseño", y la sección de "Identificación Preliminar de Riesgos" obliga a considerar la "Equidad y no discriminación" desde la concepción.

2. Valoración de Viabilidad

Descripción de la Actividad: Una vez que el Sponsor de Negocio tiene un borrador del canvas, este no avanza en solitario. Se comparte con un grupo de roles clave para una primera evaluación o "filtro de viabilidad":

- El Responsable Técnico evalúa si la idea es técnicamente factible con la tecnología y los datos disponibles o alcanzables.
- El Área Jurídica y de Planeación revisan la conformidad legal preliminar y la viabilidad presupuestal.

Referencias clave (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2):** Esta actividad representa la primera interacción entre los roles clave definidos en el modelo.

3. Identificación de Riesgos y Partes Interesadas

Descripción de la Actividad: Se identifican los riesgos preliminares y los actores involucrados.

- El Delegado de Protección de Datos (DPO) realiza una valoración inicial de los riesgos de privacidad.
- Se identifican los stakeholders y partes interesadas.

Referencias clave (Capítulo 4.1):

Modelo de Gobierno (Sección 4.1.2): Esta actividad representa la primera interacción entre los roles clave definidos en el modelo: Sponsor de Negocio, Responsable Técnico y miembros del Comité de IA como el DPO y los representantes de Jurídica y Planeación.

- **Matriz RACI (Sección 4.1.2.3):** La matriz para la fase "Intake y Canvas" se materializa aquí. El Sponsor es "Accountable" (A), el Responsable Técnico es "Responsible" (R) de su parte, y el Comité y Jurídica son "Consulted" (C).
- **Políticas Fundamentales (Sección 4.1.1):** La participación del DPO pone en marcha la "Política de Gobierno de Datos" (4.1.1.1), mientras que la identificación de riesgos activa la "Política de Gestión de Riesgos de IA" (4.1.1.2).

Punto de Control (Gate 1): Revisión y Decisión del Comité de IA

Descripción: El Comité de IA se reúne para revisar el AI Use-Case Canvas ya enriquecido con la valoración preliminar. Esta es la primera decisión formal de gobernanza del ciclo de vida.

Decisión: El Comité evalúa la propuesta contra cuatro criterios clave:

- **Alineación Estratégica:** ¿Responde el caso de uso a los objetivos de la entidad?
- **Viabilidad Preliminar:** ¿Es técnica, legal y presupuestalmente realista?
- **Claridad del Propósito:** ¿Está bien definido el problema y el valor que se espera generar?
- **Disponibilidad de Recursos:** ¿Existen los recursos o un plan para obtenerlos?

Basado en esto, el Comité decide si aprueba el paso a la Fase 2 (Clasificación de Riesgo), lo rechaza, o lo devuelve para que se aclare o complete la información.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2.1):** Este es el primer acto de supervisión del Comité de IA, que actúa como el guardián estratégico del portafolio de proyectos de IA.
- **Métricas y KPIs (Sección 4.1.5):** La sección "Métricas de Éxito" del canvas es revisada por el comité para asegurar que el proyecto busca generar un valor público medible y está alineado con las métricas de impacto institucional.

Fase 2 - Clasificación de Riesgo

Descripción: Se asigna un nivel de riesgo (inaceptable, alto, limitado o mínimo).

Actividades:

- Evaluación de matriz de clasificación.
- Asignación de nivel de riesgo.
- Activación de obligaciones reforzadas para alto riesgo.

Gate de Revisión G2 - Comité de IA: El Comité de IA revisa y valida la clasificación.

Entregable: Ficha de Clasificación de Riesgo.

Herramientas: Matriz de Riesgo

Guía de Implementación: Fase 2 - Clasificación de Riesgo

Esta fase se activa una vez que el AI Use-Case Canvas ha sido aprobado en el Gate 1. Es una fase corta pero de máximo impacto estratégico.

Matriz RACI para esta fase:

- **Comité de IA:** Accountable (A)
- **DPO:** Consultado (C)
- **Responsable Técnico:** Responsible (R)
- **Sponsor de Negocio:** Informed (I)
- **Área Jurídica:** Consultado (C)

Actividades Clave y sus Referencias en el Framework

1. Evaluación de Matriz de Clasificación

Descripción de la Actividad: El equipo del proyecto, liderado por el Sponsor de Negocio y el Responsable Técnico, evalúa el caso de uso contra los criterios definidos en la política de riesgos. Se debe analizar:

- **Propósito del sistema:** ¿Para qué se usará? (ej. evaluar acceso a un servicio, optimizar una ruta).
- **Población afectada y tipo de decisión:** ¿Afecta derechos fundamentales o el acceso a servicios esenciales?
- **Datos procesados:** ¿Utiliza datos sensibles o biométricos?
- **Consecuencias de errores:** ¿Cuál es el impacto potencial de un fallo?

Referencias clave (Capítulo 4.1):

- **Política de Gestión de Riesgos de IA (Sección 4.1.1.2):** Esta actividad es la aplicación directa del "Sistema de clasificación de riesgos" descrito en dicha política. Las definiciones de cada nivel de riesgo (ej. qué se considera "inaceptable" o "alto riesgo") provienen de esta sección.
- **Principios Fundamentales (Sección 4.1.1):** La clasificación es la primera materialización concreta de los principios. Un sistema que viole el "Respeto por los derechos humanos" se clasificará como Inaceptable. Uno que afecte la "Equidad" en el acceso a programas sociales será de Alto Riesgo. Un chatbot que solo requiera ser transparente sobre su naturaleza será de Riesgo Limitado, cumpliendo el principio de "Transparencia".

2. Asignación de Nivel de Riesgo

Descripción de la Actividad: Basado en el análisis anterior, el sistema se clasifica en una de las cuatro categorías: Inaceptable, Alto, Limitado o Mínimo.

3. Activación de Obligaciones Reforzadas

Descripción de la Actividad: Si es Alto Riesgo, se activan las "obligaciones reforzadas". Esto significa que el proyecto deberá cumplir con los requisitos más estrictos en las fases siguientes:

- **Si es Riesgo Inaceptable:** El proceso se detiene por completo y el proyecto se cancela. No puede avanzar bajo ninguna circunstancia.

- **Si es Alto Riesgo:** Se activan las "obligaciones reforzadas". Esto significa que el proyecto deberá cumplir con los requisitos más estrictos en las fases siguientes: el ARA/DPIA de la Fase 3 se vuelve obligatorio, se exigirá documentación técnica exhaustiva en la Fase 5, las pruebas de la Fase 6 serán más rigurosas y las auditorías de la Fase 8 deberán ser más frecuentes e incluir una evaluación externa.
- **Si es Riesgo Limitado o Mínimo:** El camino es más ágil. La Fase 3 (ARA/DPIA) puede no ser necesaria (a menos que se traten datos sensibles), y los requisitos de documentación y auditoría son menos intensivos.

Referencias clave (Capítulo 4.1):

- **Ciclo de Vida (Sección 4.1.3):** La clasificación en esta fase es un "interruptor" que modifica el flujo del ciclo de vida. Determina si la Fase 3 (ARA/DPIA) es obligatoria o no.
- **Controles Clave (Sección 4.1.4):** La profundidad con la que se implementarán y auditarán los controles de "Supervisión humana significativa" (4.1.4.1), "Documentación técnica" (4.1.4.4) y otros, es proporcional al nivel de riesgo asignado aquí.

Punto de Control (Gate 2): Validación de la Clasificación

Descripción: El Comité de IA revisa la clasificación propuesta por el equipo y la justificación que la sustenta. No es una simple formalidad; el Comité debe estar convencido de que la clasificación es correcta, ya que esto tiene implicaciones significativas en los recursos y el tiempo que consumirá el proyecto.

Decisión: El Comité valida formalmente la clasificación. Esta decisión se documenta en la Ficha de Clasificación de Riesgo.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2):** La Matriz RACI (4.1.2.3) establece que el Comité de IA (4.1.2.1) es "Accountable" (A) de esta decisión. Es una de sus responsabilidades más importantes.
- **Instrumentos Operativos (Sección 4.1.2.4):** El resultado de esta fase (el nivel de riesgo) se convierte en un campo clave y obligatorio en el "Registro Central de Casos de Uso de IA", permitiendo tener una visión global del portafolio de riesgos de IA de la entidad.

Fase 3 - ARA/DPIA (alto riesgo o datos sensibles)

Descripción: Identificar y mitigar impactos sobre derechos fundamentales y la privacidad.

Actividades:

- Completar Plantilla ARA/DPIA.
- Análisis de riesgos.
- Propuestas de medidas de mitigación.

Gate de Revisión G3 - DPO: El DPO aprueba o rechaza el documento.

Entregable: ARA/DPIA aprobado con plan de mitigación.

Herramientas: Plantilla ARA-DPIA

Guía de Implementación: Fase 3 - ARA/DPIA

Esta fase es el ejercicio de debida diligencia por excelencia para los sistemas más críticos. Es liderada por el Delegado de Protección de Datos (DPO) en estrecha colaboración con los responsables técnico y de negocio.

Matriz RACI para esta fase:

- **Comité de IA:** Accountable (A)
- **DPO:** Accountable (A)
- **Responsable Técnico:** Consultado (C)
- **Sponsor de Negocio:** Consultado (C)
- **Área Jurídica:** Consultado (C)

Actividades Clave y sus Referencias en el Framework

1. Completar Plantilla ARA/DPIA

Descripción de la Actividad: Esta actividad consiste en completar de manera exhaustiva la Plantilla ARA/DPIA. No es un simple checklist, sino una investigación detallada que incluye:

- **Mapeo de Datos:** Describir en detalle el flujo de los datos personales, desde su recolección hasta su eliminación.
- **Análisis de Necesidad y Proporcionalidad:** Justificar por qué es necesario usar los datos y el sistema de IA, y por qué es proporcional al problema que se quiere resolver.

- **Identificación de Riesgos Específicos:** Utilizando la Matriz de Riesgos de IA, se identifican y califican los riesgos específicos para los derechos y libertades, como la privacidad, la no discriminación, el debido proceso y la equidad.

2. Análisis de Riesgos

Descripción de la Actividad: Utilizando la Matriz de Riesgos de IA, se identifican y califican los riesgos específicos para los derechos y libertades, como la privacidad, la no discriminación, el debido proceso y la equidad.

Referencias clave (Capítulo 4.1 y 4.2):

- **Caja de Herramientas (Sección 4.2):** Esta actividad es la aplicación directa de dos herramientas clave: la "Plantilla ARA / DPIA" (4.2.3), que estructura el análisis, y la "Matriz de Riesgos de IA" (4.2.2), que se usa dentro de la plantilla para evaluar la probabilidad e impacto de cada riesgo identificado.
- **Política de Gobierno de Datos (Sección 4.1.1.1):** Es la implementación formal del requisito de "Evaluaciones de impacto en privacidad", alineado con las directrices de la SIC (Circular Externa 002/2024).
- **Política de Gestión de Riesgos de IA (Sección 4.1.1.2):** Se aplican los pasos de "Identificación de riesgos" y "Medición de riesgos" de la metodología basada en el NIST AI RMF que se describe en la política.

3. Propuestas de Medidas de Mitigación

Descripción de la Actividad: El resultado del análisis no es solo una lista de riesgos, sino un plan de acción concreto para gestionarlos. Para cada riesgo significativo identificado, se debe proponer una o más medidas de mitigación.

- **Técnicas:** "El sistema deberá usar seudonimización en la base de datos de producción".
- **Organizativas:** "Se deberá implementar un protocolo de supervisión humana de tipo 'human-in-the-loop' para todas las decisiones que rechacen una solicitud".
- **De Validación:** "En la fase de pruebas, se deberá demostrar un Disparate Impact Ratio superior a 0.9 para la variable de género".

Referencias clave (Capítulo 4.1):

- **Controles Clave (Sección 4.1.4):** El plan de mitigación es, en esencia, una selección personalizada de los controles definidos en esta capa. El equipo del proyecto elige, a partir

de este catálogo de controles, cuáles son indispensables para hacer que el riesgo del sistema sea aceptable. El plan debe especificar QUÉ control se aplicará (ej. "No discriminación y equidad"), CÓMO se implementará y CÓMO se medirá su eficacia.

- **Política de Gestión de Riesgos de IA (Sección 4.1.1.2):** Corresponde al paso de "Gestión de riesgos", que consiste en seleccionar y aplicar medidas proporcionales para tratar los riesgos.

Punto de Control (Gate 3): Aprobación Vinculante del DPO

Descripción: Este es uno de los puntos de control más rigurosos. El documento ARA/DPIA finalizado, junto con su plan de mitigación, se presenta al Comité de IA para su aprobación. En esta decisión, el DPO tiene un voto decisivo o vinculante.

Decisión:

- Si el DPO considera que los riesgos para la privacidad no están adecuadamente mitigados, puede vetar el proyecto, deteniéndolo hasta que se propongan soluciones satisfactorias.
- El Comité, con el visto bueno del DPO, aprueba formalmente el documento. Esta aprobación convierte el plan de mitigación en un "contrato" interno que el equipo técnico debe cumplir en las fases de desarrollo y pruebas.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2):** La Matriz RACI (4.1.2.3) otorga al DPO un poder singular en esta fase, marcándolo como "Accountable" (A) y con voto vinculante. Esto subraya la prioridad que el framework le da a la protección de datos en sistemas de alto riesgo. El Comité de IA también es "Accountable", ratificando la decisión.
- **Ciclo de Vida (Sección 4.1.3):** El plan de mitigación aprobado aquí se convierte en la principal fuente de requisitos no funcionales para la Fase 5 (Desarrollo) y define los criterios de éxito para la Fase 6 (Pruebas).

Fase 4 - Gobierno de Datos

Descripción: Asegurar que los datos sean de alta calidad, representativos y gestionados éticamente.

Actividades:

- Inventario y documentación de fuentes.
- Evaluación de calidad dimensional.

- Análisis de representatividad y detección de sesgos.
- Elaboración de Data Sheets.

Gate de Revisión G4 - Propietario de datos y DPO: Certifican la calidad y la gobernanza adecuada.

Entregable: Data Sheets aprobados y plan de gobernanza de datos.

Herramientas: DataSheet

Guía de Implementación: Fase 4 - Gobierno de Datos

Esta fase se centra en el trabajo práctico y profundo con los conjuntos de datos. Es liderada por el Propietario de Datos (Data Steward) y el Responsable Técnico, con la supervisión clave del DPO.

Matriz RACI para esta fase:

- **Comité de IA:** Consultado (C)
- **DPO:** Accountable (A)
- **Responsable Técnico:** Responsible (R)
- **Sponsor de Negocio:** Consultado (C)
- **Área Jurídica:** Informed (I)

Actividades Clave y sus Referencias en el Framework

1. Inventario y Documentación de Fuentes

Descripción de la Actividad: Se identifican y documentan todas las fuentes de datos (internas, externas, terceros).

2. Evaluación de Calidad Dimensional

Descripción de la Actividad: Se evalúa su calidad dimensional utilizando una lista de verificación para medir completitud (valores faltantes), precisión (errores), consistencia y actualidad.

Referencias clave (Capítulo 4.1):

- **Política de Gobierno de Datos (Sección 4.1.1.1):** Esta actividad es la implementación directa de esta política, específicamente de los componentes "Calidad y representatividad de datos" y "Gestión de sesgos y discriminación".

- **Métricas y KPIs (Sección 4.1.5.2):** Aquí se establecen las líneas base para las "Métricas de Calidad de Datos". Se mide por primera vez la "Completitud de datos" y se evalúa la "Representatividad demográfica" para ver si cumple con los umbrales aceptables (ej. p-valor ≥ 0.05).

3. Análisis de Representatividad y Detección de Sesgos

Descripción de la Actividad: Se realiza un análisis estadístico para determinar si la distribución demográfica del conjunto de datos refleja fielmente a la población objetivo. Cuando sea técnica y éticamente viable, se aplican técnicas para mitigar los sesgos encontrados.

4. Elaboración de Data Sheets

Descripción de la Actividad: Usando la plantilla del toolkit, se crea un "datasheet" para cada conjunto de datos. Este documento es una "hoja de vida" del dataset que documenta de forma transparente: su origen, metodología de recolección, composición, procesos de limpieza y pre-procesamiento aplicados, y, de manera crucial, los sesgos conocidos y las limitaciones que no pudieron ser eliminadas.

Referencias clave (Capítulo 4.1 y 4.2):

- **Principio "Equidad y no discriminación" y "Transparencia" (Sección 4.1.1):** La mitigación de sesgos atiende directamente a la equidad. La creación del Data Sheet es un acto de transparencia fundamental, reconociendo que ningún conjunto de datos es perfecto.
- **Caja de Herramientas (Sección 4.2.4):** La actividad se centra en el uso y diligenciamiento de la plantilla "DataSheet", inspirada en la propuesta de Gebru et al. (2018).
- **Controles Clave (Sección 4.1.4):** Se implementa el control de "No discriminación y equidad" (4.1.4.1) que exige el análisis y mitigación de sesgos, y el control de "Documentación técnica" (4.1.4.4), del cual el Data Sheet es un componente esencial.

Punto de Control (Gate 4): Certificación de la Calidad de los Datos

Descripción: El Propietario de Datos (como responsable del activo de información) y el DPO (como garante de la privacidad) realizan la revisión final de los Data Sheets y el plan de gobernanza.

Decisión: Ambos roles deben certificar que los datos tienen la calidad suficiente, los riesgos de privacidad están gestionados y los sesgos y limitaciones están documentados de forma transparente y son aceptables para el caso de uso. Para sistemas de alto riesgo, se podría exigir una validación estadística formal por parte de un tercero antes de dar la aprobación. La aprobación significa que los datos están "listos para ser usados" en el desarrollo del modelo.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2):** Este gate empodera al rol de "Propietario de Datos (Data Steward)" (4.1.2.2). La Matriz RACI (4.1.2.3) confirma que el DPO es "Accountable" (A) y el Responsable Técnico es "Responsible" (R) de esta fase.
- **Ciclo de Vida (Sección 4.1.3):** Los Data Sheets aprobados se convierten en un insumo indispensable para la Fase 5 (Desarrollo), ya que le dicen al equipo de desarrollo exactamente con qué datos pueden trabajar, y para la Fase 6 (Pruebas), que usará este mismo conjunto de datos validado para evaluar el modelo.

Fase 5 - Desarrollo o Adquisición

Descripción: Traducir los requisitos de gobernanza, seguridad y calidad en un producto técnico.

Actividades: Desarrollo interno o adquisición externa.

Gate de Revisión G5: Comité técnico o Comité de IA, según la vía.

Entregable: Sistema con documentación técnica o contrato firmado.

Herramientas: Checklist de Proveedores

Guía de Implementación: Fase 5 - Desarrollo o Adquisición

Esta fase es ejecutada por el equipo técnico (para desarrollo) o por un equipo conjunto de contratación, jurídico y técnico (para adquisición).

Matriz RACI para esta fase:

- **Comité de IA:** Consultado (C)
- **DPO:** Consultado (C)
- **Responsable Técnico:** Accountable (A)
- **Sponsor de Negocio:** Consultado (C)
- **Área Jurídica:** Accountable (A) - para adquisiciones

Camino A: Desarrollo Interno

Si la entidad decide construir la solución con sus propios recursos.

Actividad: Diseño e Implementación con Gobernanza Integrada

Descripción de la Actividad: El equipo de desarrollo utiliza los entregables de las fases anteriores como su biblia de requisitos no funcionales.

- **Implementación de Controles:** El plan de mitigación del ARA/DPIA se convierte en historias de usuario o tareas técnicas. Por ejemplo, si el plan exige seudonimización, el equipo debe implementar esa función. Si exige un mecanismo de supervisión humana, deben diseñar y codificar esa interfaz.
- **Privacidad y Seguridad por Diseño:** Se aplican desde el inicio los principios de minimización de datos (usando solo los datos definidos en los Data Sheets), controles de acceso y cifrado. Se implementan defensas contra ataques adversariales comunes.
- **Trazabilidad y Versionamiento:** Se debe utilizar un sistema de control de versiones (como Git) para el código, los modelos entrenados y los datos de entrenamiento. Esto es crucial para la reproducibilidad y la auditoría.
- **Documentación Continua:** Se documentan las decisiones de diseño, los hiperparámetros del modelo y se empieza a construir la Model Card.

Referencias clave (Capítulo 4.1):

- **Controles Clave (Sección 4.1.4):** Esta actividad es la implementación directa en código de los controles seleccionados en el plan de mitigación, especialmente los de "Privacidad y Protección de Datos" (4.1.4.2) y "Seguridad y Resiliencia" (4.1.4.3).
- **Principio "Privacidad por diseño y por defecto" (Sección 4.1.1):** Se materializa este principio en la arquitectura del software.
- **Entregables de Fases Anteriores:** El ARA/DPIA (Fase 3) y los Data Sheets (Fase 4) son los documentos de requisitos principales para el equipo de desarrollo.

Camino B: Adquisición Externa

Si la entidad decide comprar una solución a un proveedor.

Actividad 1: Debida Diligencia y Selección del Proveedor

Descripción de la Actividad: Se evalúa a los potenciales proveedores no solo por el precio o las funcionalidades, sino por su madurez en IA responsable.

- **Uso del Checklist de Proveedores:** Se aplica esta herramienta para evaluar la conformidad regulatoria del proveedor, su capacidad para entregar documentación (Model Cards, Data Sheets), sus certificaciones de seguridad y su experiencia.
- **Consulta del Catálogo de Proveedores:** Se debe priorizar a los proveedores que ya figuren en el "Catálogo Distrital de Proveedores Pre-evaluados" para agilizar el proceso.

Referencias clave (Capítulo 4.1 y 4.2):

- **Política de Compras y Proveedores de IA (Sección 4.1.1.3):** Esta actividad es la aplicación directa de toda esta política.
- **Caja de Herramientas (Sección 4.2.5):** La herramienta central de esta actividad es el "Checklist de Evaluación de Proveedores de IA".
- **Instrumentos Operativos (Sección 4.1.2.4):** Se debe consultar y hacer uso del "Catálogo Distrital de Proveedores Pre-evaluados".

Actividad 2: Negociación del Contrato con Cláusulas de Gobernanza

Descripción de la Actividad: El equipo jurídico, el DPO y el equipo técnico deben asegurar que el contrato contenga cláusulas específicas y exigibles que protejan a la entidad.

- **Cláusulas de Datos:** Incluir un Acuerdo de Procesamiento de Datos (DPA) que defina roles y responsabilidades según la Ley 1581.
- **Cláusulas de Auditoría y Documentación:** Exigir por contrato el derecho de la entidad a auditar el sistema y la obligación del proveedor de entregar Model Cards y Data Sheets completos y actualizados.
- **Cláusulas de Seguridad y SLA:** Definir Acuerdos de Nivel de Servicio (SLAs) para rendimiento y disponibilidad, y obligaciones claras de notificación en caso de incidentes de seguridad.

Referencias clave (Capítulo 4.1):

- **Política de Compras y Proveedores de IA (Sección 4.1.1.3):** Se implementan los requisitos contractuales definidos en esta política.
- **Control Clave 4.1.4.6 "Gestión de Terceros (Proveedores)":** Se materializa el control de "Cláusulas contractuales obligatorias".

Punto de Control (Gate 5): Revisión de la Solución o Contrato

Descripción: La revisión depende del camino elegido:

- **Desarrollo Interno:** Un comité técnico revisa la arquitectura y el código para verificar que se hayan implementado los controles de gobernanza.
- **Adquisición Externa:** El Comité de IA, el DPO y Jurídica revisan el contrato y la documentación del proveedor para confirmar que cumplen con todo lo exigido en el checklist y el ARA/DPIA.

Decisión: Se aprueba el paso a la Fase 6 (Pruebas) solo si se verifica que los requisitos de gobernanza están sólidamente integrados en la solución técnica o en el marco contractual.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2.3):** La Matriz RACI define la rendición de cuentas: el Responsable Técnico es "Accountable" (A) del desarrollo, mientras que el área Jurídica es "Accountable" (A) de los contratos. El Comité de IA y el DPO son consultados (C) y validan el cumplimiento.

Fase 6 - Pruebas y Validación

Descripción: Demostrar cumplimiento con requisitos funcionales, técnicos, éticos y de usabilidad.

Actividades: Pruebas técnicas, de equidad, de explicabilidad, de usabilidad y de integración.

Gate de Revisión G6 - Comité de IA: El Comité de IA aprueba basándose en la evidencia documental.

Entregable: Informe de pruebas y validación con Model Card.

Herramientas: ModelCard

Guía de Implementación: Fase 6 - Pruebas y Validación

Esta fase se centra en la ejecución de un plan de pruebas exhaustivo que cubra todas las dimensiones de la IA responsable.

Matriz RACI para esta fase:

- **Comité de IA:** Consultado (C)
- **DPO:** Consultado (C)
- **Responsable Técnico:** Accountable (A)

- **Sponsor de Negocio:** Consultado (C)
- **Área Jurídica:** Informed (I)

Actividades Clave y sus Referencias en el Framework

1. Pruebas Técnicas

Descripción de la Actividad: Se valida la estabilidad y confiabilidad del sistema desde una perspectiva de ingeniería. Esto incluye:

- **Rendimiento:** Medir métricas de desempeño del modelo (precisión, recall, F1, etc.) contra un conjunto de datos de validación independiente y nunca antes visto por el modelo.
- **Robustez:** Evaluar cómo se comporta el sistema ante datos inesperados, ruido, o pequeñas perturbaciones (ataques adversariales de evasión), asegurando una "degradación graciosa" en lugar de fallos catastróficos.
- **Seguridad:** Realizar pruebas de penetración (pentesting) para identificar y corregir vulnerabilidades de ciberseguridad.
- **Escalabilidad:** Simular cargas de trabajo para asegurar que el sistema puede manejar el volumen de peticiones esperado en producción con tiempos de respuesta aceptables.

Referencias clave (Capítulo 4.1):

- **Principio "Seguridad y robustez" (Sección 4.1.1):** Esta actividad es la validación práctica de dicho principio.
- **Control Clave 4.1.4.3 "Robustez técnica" y "Ciberseguridad":** Las pruebas implementan directamente estos controles, pasando de la teoría a la evidencia empírica.
- **Métricas 4.1.5.3 "Métricas de Desempeño Técnico":** Las pruebas generan los valores iniciales para los KPIs de precisión del modelo, disponibilidad (uptime) y tiempo de respuesta, que luego serán monitoreados en producción. Los resultados deben cumplir los umbrales definidos en el ARA/DPIA.

2. Pruebas de Equidad

Descripción de la Actividad: Esta es una de las pruebas más críticas. Su objetivo es detectar y cuantificar sesgos discriminatorios en el comportamiento del modelo.

- **Análisis Cuantitativo:** Calcular métricas de equidad (fairness) como el Disparate Impact Ratio o la Equal Opportunity Difference.
- **Resultados Desagregados:** Analizar el rendimiento del modelo (ej. tasa de error) para diferentes grupos demográficos (género, etnia, estrato, etc.) y comparar los resultados para asegurar que no haya disparidades inaceptables.
- **Documentación de Trade-offs:** Si se aplican mitigaciones, se debe documentar en la Model Card cómo estas afectan el rendimiento general del modelo (el conocido trade-off entre equidad y precisión).

Referencias clave (Capítulo 4.1):

- **Principio "Equidad y no discriminación" (Sección 4.1.1):** Estas pruebas son el mecanismo principal para garantizar el cumplimiento de este principio.
- **Política de Gobierno de Datos (Sección 4.1.1.1):** Valida la efectividad de la "Gestión de sesgos y discriminación" aplicada en la Fase 4.
- **Control Clave 4.1.4.1 "No discriminación y equidad":** Es la implementación directa del control, que exige pruebas y monitoreo de métricas de fairness.
- **Métricas 4.1.5.3 "Métricas de equidad (fairness)":** Las pruebas proporcionan los valores de los KPIs de equidad que deben cumplir con los umbrales definidos por el Comité de IA (ej. Disparate Impact Ratio no inferior a 0.8).

3. Pruebas de Explicabilidad

Descripción de la Actividad: No basta con que el sistema tenga mecanismos de explicación; estos deben ser funcionales y comprensibles.

- **Verificación Funcional:** Comprobar que las herramientas de explicabilidad (ej. LIME, SHAP) están correctamente integradas y generan explicaciones para las decisiones del modelo.
- **Validación con Usuarios:** Realizar pruebas con los perfiles de usuario definidos (ciudadanos, operadores, auditores) para asegurar que las explicaciones generadas son realmente inteligibles y útiles para ellos.

Referencias clave (Capítulo 4.1):

- **Principio "Transparencia y explicabilidad" (Sección 4.1.1):** Evalúa si el sistema es transparente en la práctica, no solo en la teoría.

- **Control Clave 4.1.4.4 "Inteligibilidad de explicaciones":** Es la validación directa de este control, asegurando que las explicaciones se adapten a sus destinatarios.

4. Pruebas de Usabilidad

Descripción de la Actividad: Se evalúa la interacción del sistema con sus usuarios.

- **Usabilidad:** Realizar sesiones con usuarios finales representativos para identificar problemas en la interfaz y el flujo de interacción.
- **Accesibilidad:** Auditar que la interfaz del sistema cumpla con los estándares de accesibilidad universal (WCAG 2.1 nivel AA) para garantizar el acceso a personas con discapacidad.

5. Pruebas de Integración

Descripción de la Actividad: Verificar que el sistema se comunica correctamente con otros sistemas legados, que las APIs funcionan según lo esperado y que los mecanismos de recuperación ante fallos están operativos.

Referencias clave (Capítulo 4.1):

- **Control Clave 4.1.4.5 "Accesibilidad universal" e "Integridad del servicio":** Estas pruebas validan que el sistema es inclusivo y no degrada la calidad del servicio.
- **Control Clave 4.1.4.1 "Acceso equitativo":** La accesibilidad es una condición indispensable para garantizar un acceso equitativo y no discriminatorio.

Punto de Control (Gate 6): Aprobación para Despliegue

Descripción: El Responsable Técnico consolida todos los resultados en el Informe de Pruebas y Validación y actualiza la Model Card con las métricas finales de desempeño, equidad y robustez. Este informe se presenta al Comité de IA.

Decisión: El Comité revisa la evidencia. Para sistemas de alto riesgo, la aprobación es estricta y depende de que se hayan cumplido todos los umbrales predefinidos. Si las pruebas revelan problemas significativos, el sistema debe regresar a la Fase 5 (Desarrollo) para su corrección.

Referencia en el Framework (Capítulo 4.1):

- **Matriz RACI (Sección 4.1.2.3):** Define al Responsable Técnico como "Accountable" (A) de la ejecución y reporte de las pruebas.

- **Caja de Herramientas (Sección 4.2):** La Model Card (4.2.6) es el entregable clave que resume los hallazgos. Es un requisito de la "Política de Compras y Proveedores" (4.1.1.3) y del control de "Documentación técnica" (4.1.4.4). La aprobación en este gate se basa en la evidencia contenida en este artefacto.

Fase 7 - Despliegue

Descripción: Poner el sistema en producción de forma controlada.

Actividades:

- Capacitación de usuarios.
- Configuración de controles.
- Despliegue gradual.
- Establecimiento de canales de reporte.
- Comunicación transparente.

Gate de Revisión G7 - Comité de IA y Sponsor de Negocio: Verifican la preparación operativa y autorizan el go-live.

Entregable: Sistema en producción con controles activos.

Herramientas: Guia de Uso Interno IA

Guía de Implementación: Fase 7 - Despliegue

Esta fase se activa después de que el sistema ha superado todas las pruebas y validaciones de la Fase 6. El entregable principal es la autorización formal de puesta en producción (go-live), resultando en un sistema operando con sus controles activos y personal capacitado.

Matriz RACI para esta fase:

- **Comité de IA:** Accountable (A)
- **DPO:** Consultado (C)
- **Responsable Técnico:** Responsible (R)
- **Sponsor de Negocio:** Consultado (C)
- **Área Jurídica:** Informed (I)

Actividades Clave y sus Referencias en el Framework

1. Capacitación de Usuarios Finales y Supervisores

Descripción de la Actividad: Se debe ejecutar un programa de formación integral para todos los usuarios que interactuarán con el sistema. Esto incluye:

- **Funcionarios/Operadores:** Capacitación sobre el funcionamiento del sistema, sus capacidades, limitaciones conocidas (documentadas en la Model Card), y los procedimientos específicos para ejercer una supervisión humana significativa. Deben saber cómo intervenir, corregir o anular decisiones algorítmicas.
- **Ciudadanos:** Cuando aplique, se deben crear guías, tutoriales o infografías que expliquen de manera sencilla cómo usar el sistema, qué esperar de él y cómo ejercer sus derechos.

Referencias clave (Capítulo 4.1):

- **Principio de "Rendición de cuentas y supervisión humana" (Sección 4.1.1):** La capacitación es un prerrequisito para que la supervisión humana sea "efectiva" y no meramente ceremonial.
- **Control Clave 4.1.4.1 "Supervisión humana significativa":** Este control exige que el personal supervisor sea competente, lo cual se logra únicamente a través de una formación adecuada.
- **Métrica 4.1.5 "Cobertura de capacitación obligatoria en IA Responsable y Generativa":** El despliegue no debe aprobarse si el personal clave no ha completado y certificado esta formación. El cumplimiento de esta métrica es un indicador de madurez institucional (Sección 4.1.6).

2. Configuración de Controles Operativos y Monitoreo

Descripción de la Actividad: Antes del go-live, el responsable técnico debe configurar y activar toda la infraestructura de monitoreo. Esto incluye:

- **Logs de Auditoría y Telemetría:** Implementar registros detallados de las operaciones del sistema, decisiones tomadas y acciones de los supervisores.
- **Alertas y Dashboards:** Poner en marcha los tableros de control (descritos en la sección de métricas) que permitirán el seguimiento en tiempo real del desempeño, la equidad y la seguridad del sistema. Se deben configurar alertas automáticas para notificar desviaciones o incidentes.

Referencias clave (Capítulo 4.1):

- **Control Clave 4.1.4.3 "Telemetría y respuesta a incidentes":** Esta actividad es la implementación directa de dicho control, habilitando la capacidad de detectar y responder a fallos o ataques.
- **Control Clave 4.1.4.4 "Trazabilidad":** La configuración de logs es fundamental para asegurar la trazabilidad y la capacidad de auditoría posterior.
- **Capa 5: Métricas y KPIs (Sección 4.1.5):** Toda esta capa depende de la correcta configuración de la telemetría. El Dashboard Integrado (4.1.5.4) debe estar operativo, mostrando métricas de:
    - Gestión de Riesgos y Monitoreo de Incidentes (4.1.5.1).
    - Calidad de Datos y Confiabilidad (4.1.5.2).
    - Desempeño Técnico y Confiabilidad de Modelos (4.1.5.3).

3. Ejecución del Despliegue Gradual

Descripción de la Actividad: En lugar de una activación masiva, se debe seguir una estrategia de despliegue progresivo para minimizar riesgos. Las opciones incluyen:

- **Piloto Limitado:** Lanzar el sistema para un grupo reducido de usuarios o en un área geográfica acotada.
- **Escalamiento Progresivo:** Aumentar gradualmente el volumen de usuarios o transacciones que maneja el sistema, mientras se monitorea de cerca su comportamiento.

Referencias clave (Capítulo 4.1):

- **Política de Gestión de Riesgos de IA (Sección 4.1.1.2):** El despliegue gradual es una táctica fundamental de mitigación de riesgos. Permite contener el impacto de posibles fallos no detectados en la fase de pruebas.
- **Principio de "Seguridad y robustez" (Sección 4.1.1):** Esta estrategia permite validar la robustez del sistema en un entorno real pero controlado, asegurando una "degradación graciosa" si surgen problemas.

4. Establecimiento de Canales de Reporte y Apelación

Descripción de la Actividad: Se deben habilitar y comunicar los canales a través de los cuales tanto funcionarios como ciudadanos pueden reportar incidentes, presentar quejas o apelar decisiones automatizadas. Estos canales deben ser accesibles, gratuitos y garantizar una respuesta en plazos razonables.

Referencias clave (Capítulo 4.1):

- **Control Clave 4.1.4.5 "Canales de reclamación y apelación":** Esta actividad implementa directamente este control, materializando el derecho de la ciudadanía a ser escuchada y a solicitar una revisión humana.
- **Métricas 4.1.5.1 y 4.1.5.2:** La existencia de estos canales es la fuente de datos para KPIs como el "Número de incidentes de IA por tipología", "Tiempo medio de respuesta a incidentes" y la "Satisfacción ciudadana (CSAT/NPS)".

5. Comunicación Transparente a la Ciudadanía

Descripción de la Actividad: Cuando el sistema interactúa directamente con el público, se debe ejecutar el plan de comunicación definido en la Fase 1 (AI Use-Case Canvas). La comunicación debe ser proactiva, clara y sencilla, informando que se está interactuando con un sistema de IA, cuál es su propósito y cómo se protegen sus derechos.

Referencias clave (Capítulo 4.1):

- **Principio de "Transparencia y explicabilidad" (Sección 4.1.1):** Es la puesta en práctica de este principio fundamental para construir confianza pública.
- **Control Clave 4.1.4.4 "Divulgaciones ciudadanas":** Cumple con la obligación de notificar a los ciudadanos sobre el uso de IA.
- **Instrumento 4.1.2.4 "Registro Central de Casos de Uso de IA":** La información de este registro puede ser la base para una comunicación pública más amplia sobre el portafolio de IA de la entidad.

Punto de Control (Gate 7): Aprobación del Go-Live

Descripción: El Comité de IA realiza la última verificación formal. El Responsable Técnico debe presentar evidencia de que todas las actividades anteriores se han completado: el personal está capacitado, los dashboards de monitoreo están activos, los canales de queja funcionan y el plan de comunicación está en marcha.

Decisión: Con la venia del Comité de IA, el Sponsor de Negocio (dueño del proceso) da la autorización final para el go-live.

Referencia en el Framework (Capítulo 4.1):

- **Matriz de Responsabilidades (RACI) (Sección 4.1.2.3):** La matriz define claramente que el Comité de IA es "Accountable" (A) para la fase de Despliegue, consolidando su rol como supervisor final antes de la puesta en producción.

Fase 8 - Monitoreo y Auditoría

Descripción: Supervisión continua para detectar desviaciones y aplicar correcciones.

Actividades: Monitoreo técnico, gestión de incidentes, auditoría periódica y gestión de cambios.

Gate de Revisión G8 - Comité de IA: Revisa trimestralmente KPIs, incidentes y auditorías.

Entregable: Reportes de monitoreo, registros de incidentes e informes de auditoría.

Herramientas: Guia de Uso Interno IA

Guía de Implementación: Fase 8 - Monitoreo y Auditoría

Esta fase comienza inmediatamente después del despliegue y se extiende durante toda la vida útil del sistema. Es la operacionalización de la vigilancia y la rendición de cuentas.

Matriz RACI para esta fase:

- **Comité de IA:** Accountable (A)
- **DPO:** Consultado (C)
- **Responsable Técnico:** Responsible (R)
- **Sponsor de Negocio:** Informed (I)
- **Área Jurídica:** Consultado (C)

Actividades Clave y sus Referencias en el Framework

1. Monitoreo Técnico

Descripción de la Actividad: Es la vigilancia automatizada y constante del sistema. Utilizando los dashboards y alertas configurados en la Fase 7, el Responsable Técnico debe supervisar activamente:

- **Métricas de Desempeño y Equidad:** Seguimiento en tiempo real de la precisión, el rendimiento y, crucialmente, las métricas de equidad para asegurar que no se degraden.
- **Detección de Drift:** Implementar herramientas que alerten sobre data drift (cambios en la distribución de los datos de entrada) o concept drift (cambios en la relación entre los datos y el resultado esperado). El drift es una de las principales causas de la degradación del rendimiento y la aparición de sesgos en producción.
- **Vigilancia de Seguridad:** Monitorear los logs en busca de patrones que sugieran intentos de ataque o explotación de vulnerabilidades.

Referencias clave (Capítulo 4.1):

- **Principio "Seguridad y robustez" y "Equidad y no discriminación" (Sección 4.1.1):** El monitoreo asegura que estos principios se mantengan en el tiempo, no solo en el momento del lanzamiento.
- **Control Clave 4.1.4.3 "Robustez técnica":** La detección de drift es una implementación directa de este control para asegurar un funcionamiento confiable.
- **Capa 5: Métricas y KPIs (Sección 4.1.5):** Esta actividad es el motor que alimenta el Dashboard de Gobernanza (4.1.5.4). Proporciona los datos para KPIs críticos como la "Tasa de detección de drift" (4.1.5.2) y todas las métricas de "Desempeño Técnico y Equidad" (4.1.5.3).

2. Gestión de Incidentes

Descripción de la Actividad: Implica tener un proceso formal para manejar cualquier evento no deseado. El ciclo de vida de un incidente es:

- **Registro y Clasificación:** Todo incidente (sea un fallo técnico, una queja ciudadana, una alerta de seguridad o un problema de sesgo) debe ser registrado y clasificado según su tipología y severidad.
- **Respuesta y Remediación:** Activar los protocolos definidos para contener el problema, investigarlo, aplicar una solución y verificar que esta haya sido efectiva.
- **Comunicación:** Informar a las partes interesadas, incluyendo a los ciudadanos afectados o a las autoridades (como la SIC en caso de una brecha de datos personales), según lo exija la normativa.

Referencias clave (Capítulo 4.1):

- **Control Clave 4.1.4.3 "Telemetría y respuesta a incidentes":** Es la ejecución del protocolo de respuesta definido en este control.
- **Control Clave 4.1.4.5 "Canales de reclamación y apelación":** Este proceso gestiona las entradas recibidas a través de dichos canales.
- **Métricas 4.1.5.1 "Métricas de Gestión de Riesgos":** Esta actividad genera los datos para los KPIs "Número de incidentes de IA por tipología" y "Tiempo medio de respuesta a incidentes", que miden la eficacia de la gestión de riesgos operativos.

3. Auditoría Periódica

Descripción de la Actividad: La auditoría es una revisión formal y estructurada para verificar el cumplimiento y la efectividad de los controles.

- **Auditorías Internas (Trimestrales):** El equipo de Control Interno, junto con el Comité de IA, debe revisar el cumplimiento de los controles del ARA/DPIA, analizar los registros de quejas y, fundamentalmente, auditar los logs del sistema para validar que la supervisión humana se está ejerciendo de manera efectiva.
- **Auditorías Externas (Anuales para Alto Riesgo):** Los sistemas de alto riesgo deben ser evaluados anualmente por un tercero independiente que verifique la conformidad técnica, ética y regulatoria del sistema.

Referencias clave (Capítulo 4.1):

- **Principio "Rendición de cuentas y supervisión humana" (Sección 4.1.1):** La auditoría es el principal mecanismo de rendición de cuentas.
- **Política de Compras (Sección 4.1.1.3):** Activa las cláusulas de "derecho a auditoría" exigidas a los proveedores.
- **Modelo de Gobierno (Sección 4.1.2):** Involucra directamente al "Comité de IA", al "Representante de Control Interno" y al "DPO" en sus roles de supervisión, tal como lo define la matriz RACI (4.1.2.3).

4. Gestión de Cambios

Descripción de la Actividad: Ningún sistema de IA es estático. Cuando se deba realizar una actualización (ej. reentrenar el modelo con nuevos datos, actualizar librerías o cambiar el código), se debe seguir un proceso formal para no introducir nuevos riesgos.

- **Evaluación de Impacto:** Analizar qué impacto tendrá el cambio en el rendimiento, la equidad o la seguridad del sistema.
- **Re-ejecución de Pruebas:** Un cambio significativo debe obligar a re-ejecutar las pruebas críticas de la Fase 6. No se puede asumir que el comportamiento será el mismo.
- **Actualización de Documentación:** Si el cambio es aprobado y desplegado, se deben actualizar los artefactos de gobernanza, especialmente la Model Card y los Data Sheets.

Referencias clave (Capítulo 4.1):

- **Ciclo de Vida (Sección 4.1.3):** Reconoce que la gobernanza es iterativa. La gestión de cambios representa un "mini-ciclo" que puede requerir volver a fases anteriores como la de Pruebas (Fase 6).
- **Control Clave 4.1.4.4 "Documentación técnica":** Exige explícitamente la actualización de la documentación para mantener la transparencia y la trazabilidad.

Punto de Control (Gate 8): Revisión Trimestral de Desempeño

Descripción: Trimestralmente (o con mayor frecuencia si el riesgo lo amerita), el Comité de IA se reúne para revisar el Dashboard de Gobernanza. Analizan las tendencias de los KPIs, los incidentes más relevantes y los hallazgos de las auditorías.

Decisión: Basándose en esta evidencia, el Comité toma una de tres decisiones estratégicas:

- **Mantener:** El sistema opera correctamente. Se continúa con el monitoreo.
- **Mejorar:** Se han detectado problemas o áreas de mejora. Se asignan acciones correctivas.
- **Retirar:** El sistema presenta riesgos inaceptables, es obsoleto o su costo-beneficio es desfavorable. Se inicia la Fase 9 - Retiro.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2.1):** Es la función principal del Comité de IA: supervisar el portafolio de IA y gestionar su ciclo de vida.

- **Métricas y KPIs (Sección 4.1.5.4):** La decisión del comité está informada por el "Dashboard Integrado", haciendo de la gobernanza una práctica basada en datos.
- **Modelo de Madurez (Sección 4.1.6):** Una organización que ejecuta esta fase de manera robusta y toma decisiones basadas en evidencia demuestra un Nivel de Madurez 2 (Documentado) o 3 (Optimizado).

Fase 9 - Retiro o Fin de Vida

Descripción: Retiro ordenado que preserve la continuidad del servicio y garantice la gestión de datos.

Actividades:

- Decisión de retiro.
- Planificación de la transición.
- Gestión de datos.
- Documentación de lecciones aprendidas.

Gate de Revisión G9 - Comité de IA y DPO: Aprueban el plan de retiro.

Entregable: Sistema retirado, datos gestionados e informe de lecciones aprendidas.

Guía de Implementación: Fase 9 - Retiro o Fin de Vida

Esta fase se inicia tras una decisión formal del Comité de IA, generalmente tomada durante la Fase 8 (Monitoreo), basada en la evidencia de que el sistema ya no es viable, necesario o conforme.

Matriz RACI para esta fase:

- **Comité de IA:** Accountable (A)
- **DPO:** Accountable (A)
- **Responsable Técnico:** Responsible (R)
- **Sponsor de Negocio:** Consultado (C)
- **Área Jurídica:** Consultado (C)

Actividades Clave y sus Referencias en el Framework

1. Decisión de Retiro

Descripción de la Actividad: La decisión de retirar un sistema debe estar fundamentada y dar lugar a un plan de acción detallado.

- **Triggers para la Decisión:** La decisión se basa en factores como obsolescencia técnica, cambios regulatorios que lo hacen ilegal, riesgos inaceptables detectados en el monitoreo, o un análisis costo-beneficio desfavorable.
- **Plan de Transición:** Se debe crear un plan que detalle cómo se gestionará la continuidad del servicio. Esto puede implicar el retorno a un proceso manual anterior, la migración a un nuevo sistema o simplemente la discontinuación del servicio si ya no es necesario.

Referencias clave (Capítulo 4.1):

- **Fase 8 - Monitoreo:** La decisión de retiro es una de las tres posibles salidas del Gate 8.

2. Planificación de la Transición

Descripción de la Actividad: Se debe crear un plan que detalle cómo se gestionará la continuidad del servicio. Esto puede implicar el retorno a un proceso manual anterior, la migración a un nuevo sistema o simplemente la discontinuación del servicio si ya no es necesario.

- **Modelo de Gobierno (Sección 4.1.2):** El Comité de IA (4.1.2.1) es la entidad que toma la decisión formal, cumpliendo con su rol de supervisar todo el ciclo de vida, como se estipula en la Matriz RACI (4.1.2.3).
- **Métricas 4.1.5.3 "Relación costo-beneficio":** Un ROI (Retorno de la Inversión) consistentemente negativo es un disparador cuantitativo clave para iniciar esta fase.

3. Gestión de Datos

Descripción de la Actividad: Esta es la actividad más crítica de la fase. El Propietario de Datos (Data Steward), bajo la supervisión del DPO, debe ejecutar un plan para todos los datos procesados y generados por el sistema.

- **Retención Legal:** Identificar y exportar de forma segura cualquier dato que deba ser conservado por obligaciones legales o para fines de auditoría.

- **Borrado Seguro:** Eliminar de forma permanente e irreversible todos los datos personales que ya no tengan una base legal para ser almacenados, cumpliendo con el derecho a la supresión.
- **Anonimización:** Como alternativa al borrado, se pueden anonimizar los datos si se desea conservarlos para análisis estadístico histórico, siempre que el riesgo de re-identificación sea nulo.

Referencias clave (Capítulo 4.1):

- **Principio "Privacidad por diseño y por defecto" (Sección 4.1.1):** Este principio aplica hasta el final del ciclo de vida de los datos, exigiendo un borrado seguro y conforme a la ley.
- **Política de Gobierno de Datos (Sección 4.1.1.1):** Esta actividad ejecuta los componentes de "Derechos de los titulares" (como el derecho a la supresión/olvido) y las políticas de retención.
- **Control Clave 4.1.4.2 "Privacidad y Protección de Datos":** La actividad es la implementación directa de los controles de la Ley 1581. El rol del DPO es central aquí, como lo confirma la Matriz RACI (4.1.2.3), donde es "Responsible" (R) y "Accountable" (A) en esta fase.

4. Documentación de Lecciones Aprendidas

Descripción de la Actividad: Se debe realizar un análisis post-mortem del proyecto para extraer conocimiento valioso y comunicarlo.

- **Lecciones Aprendidas:** Documentar qué funcionó bien y qué no a lo largo de todo el ciclo de vida del proyecto. Este informe es un activo invaluable para la organización, ya que ayuda a no repetir errores en futuras iniciativas de IA.
- **Comunicación:** Ejecutar un plan para informar a los usuarios y ciudadanos afectados sobre la discontinuación del servicio, las razones y las alternativas disponibles.

Referencias clave (Capítulo 4.1):

- **Modelo de Madurez Institucional (Sección 4.1.6):** La capacidad de aprender de la experiencia y mejorar continuamente es una característica de las organizaciones de Nivel 3 (Optimizado). Este informe es el mecanismo formal para la mejora continua.
- **Principio "Transparencia y explicabilidad" (Sección 4.1.1):** Ser transparente sobre el fin de un servicio es tan importante como serlo sobre su funcionamiento para mantener la confianza pública.

- **Control Clave 4.1.4.5 "Atención al Ciudadano":** Una comunicación clara durante el retiro es esencial para mantener la integridad del servicio y una buena relación con la ciudadanía.

Punto de Control (Gate 9): Aprobación Final y Cierre

Descripción: Es el cierre formal del ciclo de vida.

- **Aprobación del Plan:** El Comité de IA revisa y aprueba el plan de retiro completo.
- **Verificación del DPO:** El Delegado de Protección de Datos (DPO) realiza la verificación final y certifica que todos los datos han sido gestionados (retenidos, borrados o anonimizados) en estricto cumplimiento de la Ley 1581 y la política de datos de la entidad.

Decisión: Con la aprobación del Comité y la certificación del DPO, el sistema se marca oficialmente como "retirado" en el Registro Central de Casos de Uso de IA, y todos los artefactos (informes, actas) se archivan formalmente.

Referencia en el Framework (Capítulo 4.1):

- **Modelo de Gobierno (Sección 4.1.2):** El Comité de IA (4.1.2.1) ejerce su autoridad final al aprobar el plan, mientras que el DPO cumple su rol de garante de la protección de datos, como se destaca en la Matriz RACI (4.1.2.3). El proyecto se cierra en el Registro Central (4.1.2.4).

Estrategia de Adopción y Madurez Institucional

Para abordar la heterogeneidad de las entidades distritales, se propone un modelo de adopción progresiva que permite implementar este ciclo de vida de manera escalonada.

Modelo de Madurez de Gobernanza de IA (Niveles 0-3)

Este modelo permite a cada entidad autoevaluar su estado actual y trazar una ruta clara hacia la gobernanza plena.

Nivel	Estado	Descripción	Características Clave
0	**Inexistente / Ad-hoc**	El uso de IA es esporádico, sin control centralizado ni conciencia de riesgos.	• Sin inventario de IA. • Sin roles definidos. • Desarrollo "en las sombras" (Shadow IT).
1	**Inicial / Piloto**	Se reconoce la necesidad de gobernanza. Se aplica el framework en proyectos seleccionados.	• Comité de IA en conformación. • Inventario parcial. • Aplicación del Canvas y ARA/DPIA en pilotos. • Capacitación básica iniciada.

Nivel	Estado	Descripción	Características Clave
2	**Definido / Sistemático**	El ciclo de vida es el estándar institucional. Los procesos son repetibles y documentados.	• Comité de IA operativo y con actas. • 100% de casos nuevos pasan por los Gates. • Roles (DPO, Sponsor) activos. • Herramientas (Data Sheets, Model Cards) de uso obligatorio.
3	**Optimizado / Integrado**	Gobernanza proactiva basada en datos y mejora continua.	• Monitoreo automatizado (Dashboard en tiempo real). • Auditorías externas regulares. • Gestión de riesgos integrada con control interno. • Cultura de IA responsable consolidada.

Plan de Trabajo de Implementación (12 Meses)

Ruta sugerida para pasar del Nivel 0 al Nivel 2 en un año fiscal.

Trimestre 1 (Q1): Fundamentos y Concientización

- **Gobernanza:** Formalizar la creación del Comité de IA y designar roles clave (Sponsors, Enlaces Técnicos).
- **Inventario:** Realizar el levantamiento inicial de sistemas de IA existentes (Legacy).
- **Capacitación:** Ejecutar el plan de formación básico para directivos y equipos técnicos.
- **Piloto:** Seleccionar 1 caso de uso de bajo/medio riesgo para aplicar el ciclo de vida completo como aprendizaje.

Trimestre 2 (Q2): Estandarización de Procesos

- **Normativa:** Adoptar formalmente el manual de gobernanza y las plantillas (Canvas, ARA/DPIA).
- **Procesos:** Integrar los "Gates" de revisión en los procesos de contratación y gestión de proyectos de TI existentes.
- **Datos:** Iniciar la documentación (Data Sheets) de los datasets críticos de la entidad.
- **Ejecución:** Aplicar el framework a todos los *nuevos* proyectos de IA que inicien en este periodo.

Trimestre 3 (Q3): Instrumentación y Control

- **Herramientas:** Implementar el repositorio central de artefactos (Model Cards, evaluaciones).
- **Monitoreo:** Desplegar la primera versión del Dashboard de Gobernanza (ver sección siguiente).
- **Auditoría:** Realizar la primera auditoría interna de cumplimiento sobre el caso piloto y los nuevos proyectos.
- **Ajuste:** Refinar los umbrales de riesgo y métricas basados en la experiencia del primer semestre.

Trimestre 4 (Q4): Consolidación y Mejora

- **Cobertura:** Plan de remediación para sistemas "Legacy" (aplicar controles retroactivos donde sea viable).
- **Evaluación:** Medición formal del nivel de madurez alcanzado.
- **Transparencia:** Publicación del registro público de algoritmos y canales de retroalimentación ciudadana.
- **Planeación:** Definición del roadmap para alcanzar el Nivel 3 (automatización) el siguiente año.

Diseño del Tablero de Control (Dashboard) de Gobernanza

Para garantizar el monitoreo continuo (Fase 8) y la toma de decisiones basada en evidencia, las entidades deben implementar un tablero de control que visualice los KPIs definidos en la Capa 5 del framework.

Estructura del Dashboard

El tablero debe contar con tres vistas principales adaptadas a diferentes audiencias:

1. Vista Ejecutiva (Comité de IA y Alta Dirección)

Objetivo: Visión estratégica del portafolio y perfil de riesgo institucional.

- **KPIs Clave:**
    - **Inventario:** Total de sistemas activos por nivel de riesgo (Alto, Limitado, Mínimo).
    - **Cumplimiento:** % de proyectos con documentación completa (Canvas, ARA, Model Card).

- **Valor:** ROI estimado o ahorros generados por la implementación de IA.
- **Riesgo:** Número de riesgos materializados o incidentes críticos en el último periodo.

2. Vista de Supervisión y Cumplimiento (DPO, Jurídica, Control Interno)

Objetivo: Monitoreo de garantías de derechos y cumplimiento normativo.

- **KPIs Clave:**
  - **Privacidad:** % de sistemas con DPIA aprobado y vigente.
  - **Equidad:** Métricas de sesgo (ej. Disparate Impact Ratio) para sistemas críticos.
  - **Atención:** Volumen de PQRS relacionadas con IA y tiempo promedio de respuesta.
  - **Capacitación:** % de funcionarios certificados en IA responsable.

3. Vista Técnica y Operativa (Líderes Técnicos y de Datos)

Objetivo: Salud técnica de los modelos y calidad de datos.

- **KPIs Clave:**
  - **Desempeño:** Precisión (Accuracy/F1-Score) actual vs. línea base en pruebas.
  - **Estabilidad:** Tasa de detección de *Drift* (desviación de datos o modelo).
  - **Disponibilidad:** Uptime del servicio y latencia promedio.
  - **Calidad de Datos:** % de completitud y frescura de los datos de entrada.

Pautas de Implementación Técnica

- **Automatización:** Conectar el dashboard a los repositorios de código (Git), plataformas de ML (MLOps) y sistemas de PQRS para alimentación automática.
- **Alertas:** Configurar notificaciones automáticas cuando un KPI crítico (ej. equidad o latencia) cruce un umbral predefinido.
- **Acceso:** Garantizar controles de acceso basados en roles (RBAC) para proteger información sensible del desempeño de los modelos.

Reflexiones

Este ciclo de vida de gobernanza de IA de nueve fases proporciona un marco estructurado y exhaustivo para garantizar que los sistemas de inteligencia artificial se desarrollen, desplieguen y gestionen de

manera responsable, ética y conforme a la normativa vigente. Cada fase está diseñada para construir sobre la anterior, creando un proceso iterativo y robusto que abarca desde la concepción inicial hasta el retiro final del sistema.

La implementación exitosa de este ciclo de vida requiere:

- **Compromiso organizacional** desde los niveles más altos de la entidad
- **Colaboración multidisciplinaria** entre equipos técnicos, jurídicos, de negocio y de protección de datos
- **Documentación rigurosa** en cada etapa del proceso
- **Supervisión continua** a través de los gates de revisión y el Comité de IA
- **Adaptabilidad** para responder a cambios en la tecnología, la regulación y las necesidades de la ciudadanía

Al seguir este marco, las entidades del Distrito pueden garantizar que sus iniciativas de IA generen valor público mientras protegen los derechos fundamentales de los ciudadanos y mantienen la confianza pública en el uso de estas tecnologías transformadoras.

ANEXO E. FORMATOS

Anexo E1. Formato de IA Use-Case Canvas Diligenciado

Acta de Legalización del Canvas de Caso de Uso de IA

Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia | Versión: v1.0 | Exportado: 2025-12-19

Archivo cargado y validado con éxito.

1. Identificación del Caso de Uso

Título descriptivo del caso de uso	Sistema Automatizado de Validación y Expedición de Certificados de Residencia
Entidad y unidad organizacional responsable	Secretaría de Gobierno del Distrito Capital
Sponsor de Negocio (nombre, cargo, contacto)	Director de Atención al Ciudadano / Product Owner
Responsable Técnico (nombre, cargo, contacto)	Líder de Desarrollo e Innovación / Equipo TIC
Fecha de elaboración y versión	26/11/2025 - v1.0

2. Contexto y Propósito

Descripción del problema o necesidad	El proceso actual implica una validación manual de documentos que genera tiempos de espera de hasta 24 horas, depende de la capacidad humana limitada y horarios de oficina, y presenta riesgos de error en la subsanación que cierran el trámite injustificadamente.
Servicio, trámite o proceso al que aplica	Trámite de Expedición del Certificado de Residencia
Propósito específico del sistema de IA	Automatizar la clasificación y validación de documentos (Cédula y Recibos) mediante IA para emitir el certificado de forma inmediata y disponible 24/7.
Tipo de sistema de IA	Procesamiento de lenguaje natural
Resultados esperados (cuantificables)	Reducción del 95% en el tiempo de expedición (de 24h a minutos). Aumento del 23% en la satisfacción ciudadana (CSAT). Disponibilidad del servicio 24/7.

3. Actores Involucrados

Usuarios finales del sistema	Ciudadanos solicitantes del certificado (población general de Bogotá).
Perfil de usuarios	Heterogéneo, con niveles variados de alfabetización digital y acceso a dispositivos de diferente calidad (gama baja/alta).
Propietario de Datos	Dirección de Atención al Ciudadano
Stakeholders afectados por decisiones del sistema	Funcionarios de validación, equipo de TI, DPO (Oficial de Privacidad), ciudadanos.
Roles de gobernanza involucrados	DPO, TIC/Seguridad, Jurídica, Comité de IA

4. Datos Requeridos

Fuentes de datos	Imágenes o PDFs cargados por el ciudadano (Documento de Identidad, Recibo de Servicio Público).
Categorías de datos	Personales
Volumen estimado y frecuencia de actualización	No especificado en el informe inicial.
Evaluación preliminar de calidad y representatividad	Variable. Depende de la calidad de los documentos cargados por el ciudadano (fotos oscuras, escaneos malos).
Evaluación preliminar de privacidad	Incluye nombres, direcciones, número de documento de identidad.

5. Revisión Legal y Bases Jurídicas

Base legal para la prestación del servicio público	Ejercicio de funciones públicas y simplificación de trámites.
Base legal para tratamiento de datos personales	Función pública
Identificación de normativa específica aplicable	Ley 1581 de 2012 (Habeas Data) y Circular 002 de la SIC.
Evaluación preliminar de conformidad legal por área Jurídica	Viable bajo el estricto cumplimiento de la Ley 1581 y la Circular 002 de la SIC.

6. Clasificación Preliminar de Riesgo

Aplicación de criterios de clasificación (AI Act)	¿Afecta derechos fundamentales?, ¿Servicios esenciales?, ¿Decisiones sobre personas?
Clasificación preliminar	Riesgo Alto
Justificación de la clasificación	El sistema impacta el acceso a servicios esenciales, toma decisiones sobre personas y procesa datos personales masivamente.

7. Identificación Preliminar de Riesgos

Riesgos éticos	No identificado explícitamente, pero relacionado con la equidad y la transparencia.
Riesgos de privacidad	Exposición de datos personales si la seguridad en el manejo de los archivos temporales (PDFs) es débil.
Riesgos de seguridad	No especificado en el informe inicial.
Riesgos de equidad	Sesgo técnico (OCR) que falle más con documentos de baja calidad, discriminando a ciudadanos con menor acceso a tecnología.
Riesgos operativos	Sobrecarga del personal humano si la tasa de derivación de casos es muy alta.
Riesgos reputacionales	Percepción de que la entidad implementa un sistema que excluye a ciertos ciudadanos.
Mitigaciones preliminares identificadas	Implementación de un flujo de 'Human-in-the-loop' para casos de baja confianza, no rechazando automáticamente.

8. Métricas de Éxito e Indicadores de Impacto

KPIs de desempeño técnico	Precisión del Modelo (OCR/NLP) ≥ 98%, Tasa de Resolución Autónoma > 90%.
KPIs de impacto en servicio	Tiempo Promedio de Expedición < 3 minutos, Satisfacción Ciudadana (CSAT) > 85%.
KPIs de cumplimiento	Diferencia de tasa de error entre documentos de alta y baja calidad < 5%.
Línea base actual	Tiempo de expedición manual: hasta 24 horas.
Metas cuantificables a 6 y 12 meses	Reducir tiempo de expedición al 95% (minutos), aumentar CSAT en 23%.

9. Plan de Despliegue, Formación y Comunicación

Estrategia de implementación	Piloto
Alcance inicial y escalamiento planificado	Piloto controlado seguido de despliegue general.
Necesidades de capacitación identificadas	Capacitación a funcionarios para que pasen de validar todo a manejar solo excepciones.
Plan de comunicación y divulgación ciudadana	Informar explícitamente al ciudadano que interactúa con una IA y su derecho a revisión humana.
Cronograma estimado de implementación	No especificado en el informe inicial.

10. Monitoreo, Auditoría y Respuesta a Incidentes

Controles de monitoreo continuo propuestos	Tablero de control en tiempo real para monitorear KPIs.
Frecuencia de revisiones y auditorías	Revisión trimestral de métricas de equidad y monitoreo de 'Drift'.
Protocolo de respuesta a incidentes	Los casos que la IA no pueda validar con alta confianza (ej. <90% certeza) no serán rechazados, sino derivados a un funcionario humano.
Canales de reclamación ciudadana	Opción para solicitar revisión humana en caso de inconformidad.

11. Criterios y Plan de Fin de Vida

Condiciones que activarían el retiro del sistema	Si la tasa de error en poblaciones vulnerables supera los umbrales éticos o si cambios normativos prohíben la automatización.
Plan preliminar de gestión de datos al final de vida	Eliminación segura o archivo de los datos conforme a la ley al final del ciclo de vida.
Plan de transición a alternativas	Reactivación del proceso manual o migración a un nuevo sistema.

12. Aprobaciones y Decisión

Evaluación de viabilidad por Responsable Técnico	Viable
Evaluación de conformidad legal por Jurídica	Requiere ajustes
Evaluación de privacidad por DPO	Requiere DPIA
Evaluación presupuestal por Planeación	Requiere gestión
Decisión del Comité de IA	Aprobado con condiciones
Firma del Presidente del Comité y fecha	APROBADO PARA PASAR A FASE 2 (Clasificación y ARA/DPIA). Se condiciona el desarrollo a la presentación de un plan de mitigación de sesgos por calidad de imagen.

Legalización de Aprobación

Fecha de Legalización de esta Acta	18/12/2025

Certifico que la información del Caso de Uso de IA ha sido revisada y aprobada por el **Comité de IA** para su implementación, bajo las consideraciones y mitigaciones descritas en este documento.

Decisión Final del Comité de IA	**Aprobado con condiciones (Técnica: Viable / Jurídica: Requiere ajustes)**
Nivel de Riesgo General Asignado	Riesgo Alto
Aprobación DPO / Jurídica	Requiere DPIA
Firma y Fecha de Aprobación del Comité (JSON)	APROBADO PARA PASAR A FASE 2 (Clasificación y ARA/DPIA). Se condiciona el desarrollo a la presentación de un plan de mitigación de sesgos por calidad de imagen.

Nombre Completo del Auditor/Revisor
Javier Mauricio Rocha
Firma del Auditor

Nombre del Responsable del Caso de Uso
Líder de Desarrollo e Innovación / Equipo TIC
Firma del Responsable

Anexo E2. Formato de Matriz de Riesgo Diligenciado

Acta de Revisión y Aprobación de Matriz de Riesgos de IA

Fecha de Exportación de la Matriz: 18/12/2025, 21:14:43

Cargar Archivo de Matriz

Seleccionar archivo Ningún archivo seleccionado

Cargue el archivo JSON de la Matriz de Riesgos para generar el acta.

Matriz de Riesgos cargada para impresión. Puede imprimir ahora.

Matriz de Riesgos Identificados

ID	Riesgo / Descripción	Categoría	Dimensión de Afectación	P	I	Calificación (P x I)	Nivel de Riesgo	Controles Actuales	Efectividad	Mitigación Adicional Recomendada	Responsable	Estado
1	El sistema OCR puede tener una tasa de error superior al procesar imágenes de baja calidad (fotos oscuras, documentos arrugados), discriminando indirectamente a ciudadanos con dispositivos de gama baja.	Equidad	Equidad y no discriminación	3	3	9	Alto	Pruebas iniciales de rendimiento del modelo OCR base.	Baja	Implementar un protocolo 'Human-in-the-loop' que prohíbe el rechazo automático. Casos con confianza <90% se derivan a un funcionario. Realizar pruebas de equidad comparando la tasa de error entre documentos de alta y baja calidad (meta: diferencia <5%).	Líder Técnico / Equipo de Desarrollo	En progreso
2	Un rechazo incorrecto por parte del sistema (falso negativo) afecta directamente el acceso del ciudadano a un servicio esencial, vulnerando su derecho al debido proceso administrativo.	Cumplimiento	Derechos y libertades fundamentales	2	3	6	Medio	El sistema está diseñado para validar datos, no para interpretar la ley.	Media	El protocolo 'Human-in-the-loop' es el control principal. Adicionalmente, se debe ofrecer un canal claro y accesible para que el ciudadano solicite revisión humana de cualquier decisión.	Sponsor de Negocio / DPO	Mitigado
3	Exposición de datos personales (nombres, ID, direcciones) si la transmisión o el almacenamiento temporal de los archivos PDF/JPG no está debidamente cifrado o es vulnerable.	Privacidad	Seguridad y ciberseguridad	1	3	3	Bajo	Uso de HTTPS para la transmisión de datos.	Media	Implementar una política de retención mínima: los archivos se eliminan automáticamente tras la validación (o tras un periodo corto de auditoría). Exigir cifrado en reposo y cláusulas de confidencialidad reforzadas con el proveedor de nube.	Oficial de Seguridad de la Información (CISO)	Mitigado
4	Si el modelo se entrena mayoritariamente con recibos de un solo proveedor (ej. Enel), podría fallar sistemáticamente con formatos de otros proveedores (ej. Acueducto), afectando la calidad del servicio.	Continuidad del servicio	Continuidad y calidad del servicio	2	2	4	Medio	El dataset inicial contiene algunos ejemplos variados.	Baja	Certificar en el Data Sheet que el dataset de entrenamiento ha sido balanceado para incluir una muestra representativa de todos los proveedores y formatos de recibos relevantes en Bogotá.	Propietario de Datos / Data Steward	En progreso
5	Percepción pública de que la entidad implementa un sistema 'caja negra' que excluye a ciudadanos, generando desconfianza y daño reputacional.	Reputación	Reputación y confianza pública	2	2	4	Medio	Comunicado de prensa en el lanzamiento.	Baja	Implementar una estrategia de transparencia activa: el chatbot debe informar que es una IA, publicar una versión simplificada de la Model Card y explicar el derecho del ciudadano a la revisión humana.	Jefe de Comunicaciones / Sponsor de Negocio	Pendiente

Legalización de la Revisión y Controles

Certificamos que los riesgos y las medidas de mitigación de esta matriz han sido revisados y validados por el equipo responsable y el Comité de IA/Riesgos.

Fecha de Revisión/Auditoría	18/12/2025
Observaciones del Auditor	Los Riesgos declarados cumplen con lo mínimo requerido para parasar a la Fase 3
Aprobación Final	Aprobado con Plan de Acción

Nombre Completo del Auditor/Revisor
Javier Mauricio Rocha

Firma

(Auditoría / Gestión de Riesgos)

Nombre del Representante del Comité de IA
Maria Alejandra Rodriguez

Firma

(Comité de IA / DPO)

Anexo E3. Formato de ARA - DPIA Diligenciado

Acta de Aprobación DPIA / ARA

Evaluación de Impacto de Protección de Datos y Análisis de Riesgos Algorítmicos

Documento generado el 2025-12-19 | Versión del Formulario: 2.0

Cargar Archivo DPIA/ARA

Seleccionar archivo Ningún archivo seleccionado

Cargue el archivo JSON de la Evaluación DPIA/ARA para generar el acta.

DPIA/ARA cargado para impresión. Puede imprimir ahora.

1. Información General

Entidad	Secretaría de Gobierno del Distrito Capital
Unidad responsable	Dirección de Atención al Ciudadano
Título del sistema	Automatización de Validación Documental para Certificado de Residencia
Responsable ARA/DPIA	Oficial de Protección de Datos (DPO) y Líder Técnico
Responsable técnico	Líder de Desarrollo e Innovación / Equipo TIC
Fecha de elaboración	2025-11-26
Versión	1.0

2. Descripción del Sistema y Alcance

Descripción técnica del sistema	Sistema de IA (Chatbot con OCR y NLP) que valida documentos (Cédula y Recibos de Servicios Públicos en PDF/JPG) para reducir el tiempo de respuesta a minutos y estar disponible 24/7.
Finalidad específica	Validar la coincidencia entre la identidad del solicitante y la dirección del predio para expedir un acto administrativo (Certificado de Residencia).
Base legal	Ley 1581 de 2012 (Habeas Data), Circular 002 de la SIC, Ley 2052 de 2020 (Simplificación de trámites).
Población objetivo	Ciudadanos solicitantes del certificado en Bogotá D.C.
Casos de uso permitidos	Extraer y validar automáticamente el nombre, número de cédula y dirección a partir de documentos para autorizar la emisión del certificado.
Casos de uso no permitidos	Utilizar los datos extraídos (direcciones, consumo, estrato) para crear perfiles de capacidad de pago, scoring crediticio o cualquier otro fin distinto a la verificación de residencia.
Tipo de decisiones	recomendacion
Contexto de uso	El sistema opera 24/7. Las validaciones exitosas generan el certificado automáticamente. Las validaciones fallidas o con confianza <90% no son rechazadas, sino derivadas a un funcionario para revisión manual.

3. Datos y Origen de la Información

Tabla de Mapeo de Datos

Categoría	Tipo de dato	Fuente	Dato personal	Base de licitud	Finalidad	Técnicas	Sesgos	Observaciones
identificacion	imagen	Ciudadano (carga de archivo)	si	obligacion	Verificar la identidad del solicitante.	recopilacion	N/A	Corresponde a la Cédula de Ciudadanía/Extranjería.
ubicacion	imagen	Ciudadano (carga de archivo)	si	obligacion	Verificar la dirección de residencia.	recopilacion	Riesgo de sesgo si el modelo no reconoce formatos de recibos de todos los proveedores de servicios.	Corresponde al recibo de servicio público.
demografico	texto	Sistema (extraído del recibo)	si	interes	Dato incidental, no utilizado para la decisión.	analitica	N/A	Estrato socioeconómico. Se prohíbe su uso para perfilamiento.

Documentación de Dataset (Datasheet Simplificada)

Nombre del dataset	RESIDENCIA_BOG_VALIDATION_V1
Origen	Datos históricos de solicitudes previas anonimizadas y cargas controladas durante la fase piloto.
Motivación	Entrenar y validar modelos de OCR/NLP para extraer y verificar automáticamente datos de cédulas y recibos de servicios públicos.
Proceso de recolección	Carga directa por el ciudadano a través del portal web del trámite.
Procesos de limpieza y anonimización	Pre-procesamiento de imágenes para mejorar contraste y normalización de texto a formatos estándar antes de la validación.
Usos previstos	Exclusivamente para la validación de la coincidencia entre el solicitante y la dirección del predio para la expedición del certificado.
Usos no previstos	Crear mapas de calor de morosidad, compartir datos con terceros comerciales, evaluar capacidad de pago.

4. Base Legal y Consentimiento

Base legal principal	ejercicio-funciones
Bases legales complementarias	El tratamiento es necesario para la simplificación de trámites administrativos, una función pública de la entidad.
¿Se requiere consentimiento explícito?	si
Mecanismo de obtención de consentimiento	Aviso de privacidad al inicio de la interacción con el chatbot, informando que es una IA y solicitando aceptación para continuar.
Alternativas sin consentimiento	El ciudadano puede realizar el trámite de forma manual en los puntos de atención presencial de la entidad.

5. Evaluación de Impactos en Derechos

Derecho	Pregunta guía	Respuesta	Riesgo	Prob. (1–3)	Impacto (1–3)	Nivel
igualdad	¿Podría el sistema funcionar peor para ciudadanos con menor acceso a tecnología?	Sí. El OCR puede fallar más con fotos de baja calidad tomadas con celulares de gama baja, afectando a poblaciones vulnerables.	Discriminación indirecta y exclusión por brecha tecnológica.	3	3	alto
debido-proceso	¿Podría un error del sistema denegar injustificadamente el acceso a un servicio?	Sí, un 'falso negativo' en la validación podría bloquear la expedición del certificado, afectando el acceso del ciudadano a otros trámites o subsidios.	Denegación de acceso a un derecho por fallo algorítmico.	2	3	medio
privacidad	¿Están los datos personales protegidos contra accesos no autorizados?	Sí, pero una vulnerabilidad en el almacenamiento temporal de los PDFs podría exponer nombres, cédulas y direcciones.	Fuga de información personal.	1	3	bajo

6. Matriz de Riesgos Algorítmicos

ID	Descripción	Categoría	Prob.	Impacto	Punt. (P×I)	Nivel	Controles
R-01	El OCR falla más con imágenes de baja calidad, discriminando a ciudadanos con dispositivos de gama baja.	equidad	3	3	9	alto	Prohibición de rechazo automático. Casos con confianza <90% o baja calidad de imagen se derivan a un humano.
R-02	El modelo se entrena con un sesgo hacia recibos de un solo proveedor (ej. Enel), fallando con otros formatos.	robustez	2	2	4	medio	Exigencia de un Data Sheet que certifique un dataset de entrenamiento balanceado con todos los proveedores y formatos.
R-03	Exposición de datos personales si el almacenamiento temporal de los PDFs no es seguro.	privacidad	1	3	3	bajo	Implementación de política de retención mínima (eliminación automática tras validación) y cifrado en reposo.

7. Medidas de Mitigación y Controles

ID	Medida	Tipo	Responsable	Plazo	Estado	Indicador	Comentarios
R-01	Implementar un protocolo 'Human-in-the-loop' que prohíbe el rechazo automático y realizar pruebas de equidad comparando la tasa de error entre documentos de alta y baja calidad.	tecnico	Líder Técnico / Equipo de Desarrollo	2025-12-15	en_curso	Diferencia en tasa de error (alta vs. baja calidad) < 5%.	Medida crítica para garantizar la equidad y el acceso universal al servicio, alineada con la condición del Comité de IA.
R-02	Certificar en el Data Sheet que el dataset de entrenamiento ha sido balanceado para incluir una muestra representativa de todos los proveedores y formatos de recibos relevantes en Bogotá.	organizativo	Propietario de Datos / Data Steward	2025-11-30	implementado	Data Sheet v1.0 aprobado con sección de balanceo de datos.	Asegura que el modelo no falle sistemáticamente con formatos de recibos menos comunes.
R-03	Implementar una política de retención mínima (eliminación automática de archivos tras la validación) y exigir cifrado en reposo y en tránsito para todos los datos.	tecnico	Oficial de Seguridad de la Información (CISO)	2025-12-05	implementado	Logs de auditoría que confirman la eliminación de archivos y configuración de cifrado validada.	Control esencial para cumplir con el principio de minimización de datos de la Ley 1581.

8. Supervisión Humana

Mecanismos de intervención humana	Protocolo 'Human-in-the-loop' obligatorio. Los casos con confianza <90% o detectados como de baja calidad de imagen se derivan a una cola de revisión humana.
Criterios de escalamiento	Confianza del modelo OCR/NLP inferior al 90%. Detección automática de imagen de baja calidad (borrosa, oscura). Solicitud explícita del ciudadano.
Roles responsables	Funcionarios de la Dirección de Atención al Ciudadano, capacitados para gestionar la 'Bandeja de Excepciones' y sobreescribir la decisión de la IA.
Frecuencia de revisión de casos	tiempo-real
Procedimiento ante desacuerdo con el sistema	El ciudadano puede solicitar revisión humana a través de un botón en la interfaz. Si el trámite ya cerró, puede usar el canal PQR para impugnar la decisión.

9. Monitoreo Continuo y KPIs

Categoría	Indicador	Definición	Fórmula	Valor objetivo	Valor actual	Frecuencia	Responsable
equidad	Diferencia de Tasa de Error (DRE)	Compara la tasa de error del modelo entre documentos de alta calidad (grupo de referencia) y de baja calidad (grupo desfavorecido).	| TasaError(BajaCalidad) - TasaError(AltaCalidad) |	< 5%	3.8% (en pruebas)	mensual	Líder Técnico
rendimiento	Tasa de Resolución Autónoma	Porcentaje de trámites que el sistema valida exitosamente (confianza >90%) sin intervención humana.	(Trámites Autónomos / Total Trámites) * 100	> 90%	N/A (pre-despliegue)	semanal	Sponsor de Negocio

10. Comunicación y Transparencia

Mensajes de aviso a usuarios	Mensaje inicial: 'Está interactuando con un asistente automatizado. Tiene derecho a solicitar revisión humana'. Mensaje de derivación: 'No pudimos validar su documento automáticamente, un funcionario lo revisará'.
Documentos publicados	Versión simplificada de la Model Card y política de privacidad en el portal web de la entidad.
Canales de consulta ciudadana	Línea de atención telefónica y sección de preguntas frecuentes en el portal.
Canales de reporte de problemas	Sistema PQRSD con una categoría específica para 'Problemas con el trámite automatizado'.
Plan de divulgación	Campaña en el portal web y redes sociales de la entidad explicando el nuevo servicio digital.

11. Auditoría y Actualizaciones

Tipo	Alcance	Frecuencia	Fecha próxima	Responsable	Hallazgos	Acción
interna	Auditoría de equidad y desempeño del modelo OCR, y cumplimiento de controles de privacidad (retención de datos).	semestral	2026-05-26	Comité de IA / Oficina de Control Interno	N/A (primera auditoría programada post-despliegue).	Verificar el cumplimiento de la condición de aprobación del Comité de IA sobre la diferencia de tasa de error.

12. Aprobaciones y Decisión

Evaluaciones

Evaluación DPO	Aprobado. El DPA con el proveedor es robusto y se implementa privacidad por diseño (retención mínima). Se prohíbe el uso de datos para re-entrenamiento de modelos externos.
Evaluación responsable técnico	Viable. La tecnología OCR/NLP es madura. El protocolo de derivación a humanos mitiga los riesgos de exclusión por calidad de imagen.
Evaluación jurídica	Conforme. El sistema no toma decisiones de rechazo de forma autónoma, preservando el debido proceso. La existencia de un canal alternativo manual garantiza el acceso universal.

Decisión

Decisión del Comité de IA	aprobado-condiciones
Fecha de aprobación	2025-11-26
Condiciones o reservas	Aprobado, condicionado a realizar pruebas de equidad en Fase 6, comparando la tasa de error entre documentos de alta y baja calidad (diferencia no debe superar el 5%).
Fecha de próxima revisión	2026-02-26

Decisión y Aprobación Final

Decisión del Comité de IA	APROBADO CONDICIONES
Fecha de Aprobación	2025-11-26
Condiciones y Reservas Impuestas	Aprobado, condicionado a realizar pruebas de equidad en Fase 6, comparando la tasa de error entre documentos de alta y baja calidad (diferencia no debe superar el 5%).
Fecha Próxima Revisión	2026-02-26

Legalización y Firmas

Certificamos que esta Evaluación de Impacto ha sido revisada, completada y aprobada bajo las condiciones indicadas.

Nombre Completo del DPO / Oficial de Cumplimiento
Javier Mauricio Rocha Firma

Nombre del Representante del Comité de IA Mario Alexander Ortiz
Firma

*Nota: La **Evaluación del DPO** y las **Condiciones de Aprobación** se encuentran detalladas en la sección anterior.*

Anexo E4. Formato de Datasheet Diligenciado

Acta de Legalización de Ficha Técnica de Datos

Dataset: RESIDENCIA_BOG_VALIDATION_V1

Cargar Ficha Técnica JSON

Seleccionar archivo Ningún archivo seleccionado

Cargue el archivo JSON de la Ficha Técnica de Datos para generar el acta.

Ficha Técnica de Datos cargada para impresión. Puede imprimir ahora.

1. Información General del Dataset

Nombre del dataset	RESIDENCIA_BOG_VALIDATION_V1
Entidad propietaria	Secretaría de Gobierno del Distrito Capital
Área responsable	Dirección de Atención al Ciudadano / Data Steward
Versión	1.0
Fecha de creación	2025-11-26
Fecha de actualización	2025-11-26
Contacto institucional	datagov@gobiernobogota.gov.co

2. Descripción y Propósito

Descripción del contenido	Dataset compuesto por imágenes (PDF/JPG) de Cédulas de Ciudadanía y facturas de servicios públicos (Energía, Acueducto, Gas) para el trámite de Certificado de Residencia en Bogotá.
Finalidad del dataset	Entrenar y validar modelos de OCR/NLP para extraer y verificar automáticamente nombres, números de cédula, direcciones y fechas de recibos de servicios públicos.
Procesos o servicios públicos que soporta	Soporta el proceso automatizado de expedición del Certificado de Residencia.

3. Origen y Método de Recolección

Fuente de los datos	Datos históricos de solicitudes previas anonimizadas y cargas controladas durante la fase piloto.
Método de captura	Carga directa por el ciudadano a través del portal web del trámite.
Frecuencia de actualización	Según necesidad de re-entrenamiento (ej. por 'model drift').

4. Composición del Dataset

Número de registros	15000
Número de variables	4
Tipos de datos	Imágenes (PDF/JPG), texto extraído.
Representatividad poblacional	El dataset fue enriquecido para incluir muestras de facturas de todos los proveedores (Acueducto, Vanti, Enel) y fotografías de baja calidad para representar a ciudadanos con dispositivos de gama baja.

Diccionario de Datos

Nombre variable	Descripción	Tipo de dato	Valores permitidos / rango	¿Dato personal?	¿Sensible?	Obligatorio	Observaciones
documento_identidad	Imagen (PDF/JPG) de la Cédula de Ciudadanía o Extranjería.		Formatos PDF, JPG, PNG.	Sí	No	Sí	Fuente principal para extraer nombre y número de ID.
recibo_servicio	Imagen (PDF/JPG) del recibo de un servicio público domiciliario.		Formatos PDF, JPG, PNG. Vigencia no mayor a 60 días.	Sí	No	Sí	Fuente para extraer dirección y verificar coincidencia.

5. Presencia de Datos Personales y Sensibles

Identificación del tipo de datos personales/sensibles	Nombres, apellidos, número de documento de identidad, dirección de residencia.
Justificación	El tratamiento es necesario para la validación de identidad y residencia en un trámite administrativo solicitado por el titular del dato.
Base legal del tratamiento	Ejercicio de funciones públicas y simplificación de trámites (Ley 2052 de 2020).

6. Calidad del Dataset

Datos faltantes	Menos del 2%. Se identifican datos faltantes en campos no estructurados de los recibos. La estrategia es derivar a revisión humana.
Inconsistencias	Direcciones escritas con formatos no estándar (ej. 'Cll' vs 'Calle'). Se normalizan durante el pre-procesamiento.
Procesos de limpieza aplicados	Pre-procesamiento de imágenes para mejorar contraste y normalización de texto a formato estándar de la DIAN/Catastro antes de la validación.
Controles de calidad	Validación cruzada y etiquetado por doble entrada en la fase de curación de datos.
Validaciones aplicadas	Scripts para verificar la consistencia de formatos de fecha y la estructura de las direcciones normalizadas.

7. Evaluación de Sesgos y Representatividad

Sesgos identificados	Sesgo de representación inicial (80% facturas de Enel). Corregido al balancear el dataset con muestras de otros proveedores y formatos físicos.
Distribución de variables sensibles	Distribución balanceada de proveedores de servicios y calidad de imagen (alta/media/baja).
Riesgos para poblaciones vulnerables	Riesgo de exclusión de ciudadanos con dispositivos de gama baja si el modelo no es robusto a imágenes de baja calidad. Mitigado con el enriquecimiento del dataset.
Limitaciones y riesgos identificados	El dataset no contiene ejemplos de documentos manuscritos o con un alto grado de deterioro, los cuales serán gestionados por el flujo de excepción humana.

8. Procesamiento y Transformaciones Aplicadas

Normalización	Normalización de direcciones a un formato estándar para facilitar la comparación.
Imputación	No se realiza imputación de datos. Los casos con datos faltantes o ilegibles se derivan a un agente humano.
Balanceo	Se aplicó sobremuestreo (oversampling) de imágenes de baja calidad y de proveedores de servicios con menor representación para evitar sesgos.
Anonimización/seudonimización	Los datos de entrenamiento se basan en solicitudes históricas anonimizadas. Los datos en producción se eliminan tras la validación.

9. Riesgos Éticos y Legales

Riesgos de discriminación	Riesgo de discriminación indirecta por brecha tecnológica. Se mitiga con pruebas de equidad y el protocolo 'Human-in-the-loop'.
Riesgos de reidentificación	Bajo. Los datos se usan en un entorno controlado y se eliminan post-trámite.
Riesgos de uso indebido	Alto si no se controla. Se prohíbe explícitamente el uso de los datos para fines diferentes a la validación de residencia.
Riesgos de sesgos estructurales	El dataset puede reflejar sesgos socioeconómicos si la calidad de los documentos se correlaciona con el estrato. Se monitorea la tasa de error por estrato.
Riesgos de vulneración de derechos	El principal riesgo es la vulneración del derecho al debido proceso si hay un rechazo erróneo. Se mitiga con la intervención humana obligatoria para rechazos.
Cumplimiento normativo	El dataset y su tratamiento se alinean con la Ley 1581 de 2012 y la Circular 002 de la SIC.

10. Uso Permitido y No Permitido

Finalidad autorizada	Exclusivamente para la validación de la coincidencia entre el solicitante y la dirección del predio para la expedición del certificado.
Restricciones	Acceso restringido al modelo de IA y a los auditores humanos autorizados.
Usos indebidos	Prohibido usar los datos para evaluar capacidad de pago, crear mapas de morosidad o compartir con terceros comerciales.
Condiciones de acceso	Acceso mediante API Key con auditoría y solo desde la infraestructura del Distrito.

11. Seguridad del Dataset

Controles de seguridad	Cifrado en tránsito (TLS 1.3) y en reposo (AES-256).
Cifrado	Aplicado tanto en la base de datos de entrenamiento como en el almacenamiento temporal de producción.
Custodia	El Líder de Desarrollo e Innovación es el custodio técnico del dataset.
Políticas de acceso	Política de Mínimo Privilegio. El acceso se concede por rol y se audita trimestralmente.

12. Historial del Dataset

Versión	Fecha	Cambios realizados	Responsable
1.0	2025-11-26	Versión inicial. Se realizó un balanceo para incluir más muestras de recibos físicos y de proveedores distintos a Enel.	Equipo de Datos / Data Steward

Aprobación Final del Uso del Dataset

Aprobación del Equipo de Datos	Aprobado. El Data Steward certifica que el dataset ha sido balanceado y cumple con los requisitos de calidad.
Aprobación del Área Jurídica	Aprobado. El uso de los datos se enmarca en las funciones públicas de la entidad.
Aprobación del Comité de IA	Aprobado con la condición de demostrar estadísticamente que la diferencia de precisión entre formatos de alta y baja calidad es inferior al 5%.

Legalización y Firmas

Certificamos que esta Ficha Técnica ha sido revisada y que el dataset cumple con las políticas de calidad, privacidad y gobernanza para su uso en sistemas de IA.

Nombre Completo del Responsable del Dataset
Mario Alexander Ortiz Firma

Nombre del Representante del Comité de IA
Javier Mauricio Rocha Firma

Anexo E5. Formato de Checklist de Proveedores Diligenciado

Acta de Evaluación y Aprobación de Proveedor de IA

Evaluación del proveedor: VisionTech OCR Services | Caso de Uso: Sistema Automatizado de Validación y Expedición de Certificados de Residencia | Riesgo: ALTO | Checklist: ESTANDAR

Cargar Checklist de Evaluación JSON

Seleccionar archivo Ningún archivo seleccionado

Cargue el archivo JSON del Checklist de Evaluación de Proveedor para generar el acta.

Checklist cargado y cálculos completados exitosamente.

1. Evaluación Detallada por Criterio

ID	Criterio Evaluado	Puntuación (2=Cumple)	Evidencia Presentada
1. Ética y Conformidad Regulatoria (Peso: 25%)			
1.1 ⭐	Conformidad con marco regulatorio colombiano	CUMPLE	*Certificado de conformidad con Ley 1581 y Circular 002 SIC. Política de privacidad pública.*
1.2	Política de IA Responsable o Ética	PARCIAL	*Política de IA Responsable interna, no pública.*
1.3	Sistema de Gestión de IA (AIMS)	PARCIAL	*Documentación de procesos de gestión de IA, sin certificación formal.*
1.4	Gestión de Riesgos de IA	CUMPLE	*Proceso documentado de gestión de riesgos de IA para sus modelos base.*
2. Transparencia y Documentación Técnica (Peso: 20%)			
2.1	Model Card (Ficha del Modelo)	CUMPLE	*Entrega de Model Card del motor OCR base como requisito contractual.*
2.2	Data Sheet (Ficha de Datos)	CUMPLE	*Entrega de Data Sheet del dataset de entrenamiento del motor OCR base.*
2.3	Documentación Técnica para Auditoría	CUMPLE	*Acceso a documentación técnica detallada de la arquitectura del modelo.*
3. Privacidad y Protección de Datos (Peso: 20%)			
3.1	Claridad sobre Titularidad de Datos	CUMPLE	*Cláusula contractual que define al Distrito como titular de los datos de operación.*
3.2 ⭐	Data Processing Agreement (DPA)	CUMPLE	*Firma de Data Processing Agreement (DPA) robusto y alineado con la normativa colombiana.*
3.3	Gestión de Derechos de los Titulares	PARCIAL	*Procedimientos manuales para atender derechos de habeas data, con compromiso de plazos.*
3.4	Políticas de Retención y Borrado de Datos	CUMPLE	*Políticas de retención y borrado seguro configurables a través de la API del servicio.*
4. Seguridad y Robustez (Peso: 20%)			
4.1 ⭐	Certificaciones de Seguridad	CUMPLE	*Certificación ISO 27001 vigente presentada durante la licitación.*
4.2	Pruebas de Robustez y Seguridad de IA	PARCIAL	*Reporte de pruebas de robustez contra datos fuera de distribución.*
4.3	Respuesta a Incidentes de Seguridad	CUMPLE	*Protocolo de respuesta a incidentes con canal de comunicación directo.*
5. Auditoría y Rendición de Cuentas (Peso: 10%)			
5.1	Derecho a Auditoría	CUMPLE	*Cláusula contractual que establece el 'Derecho a Auditoría' por parte del Distrito.*
5.2	Trazabilidad de Decisiones	CUMPLE	*Sistema de logs inmutables accesibles vía API para trazabilidad de cada validación.*
6. Calidad del Servicio y Soporte (Peso: 15%)			
6.1 ⭐	Acuerdos de Nivel de Servicio (SLA)	CUMPLE	*SLA contractual con penalizaciones por degradación de la precisión por debajo del 95%.*
6.2	Soporte Técnico y Transferencia de Conocimiento	CUMPLE	*Plan de soporte prioritario y capacitación al equipo técnico del Distrito.*
6.3	Gestión de Cambios y Roadmap	PARCIAL	*Comparte roadmap de alto nivel, con notificación de cambios con 30 días de antelación.*

2. Resumen de Criterios Mandatorios

ID	Criterio	Estado de Cumplimiento
1.1	Conformidad con marco regulatorio colombiano	CUMPLIDO
3.2	Data Processing Agreement (DPA)	CUMPLIDO
4.1	Certificaciones de Seguridad	CUMPLIDO
6.1	Acuerdos de Nivel de Servicio (SLA)	CUMPLIDO

3. Puntuación Total Ponderada

Puntuación Bruta (Máx: 38)	**33/38**
Puntuación Total Ponderada (Máx: 2.0)	**8.7/10**

Dictamen Final y Aprobación

Fecha de Legalización del Acta	18/12/2025
Dictamen Final	RECOMENDADO
Resultado Criterios Mandatorios	CUMPLE
Justificación y Observaciones	El proveedor cumple con todos los criterios mandatorios y obtiene una puntuación ponderada superior a 8.0/10. Se incluyeron cláusulas contractuales robustas de auditoría y SLA. Se recomienda la selección para pasar a la fase de pruebas e integración.

Equipo Evaluador

Responsable Técnico Evaluador	Líder de Desarrollo e Innovación / Equipo TIC	**Fecha Evaluación Técnica**	26/11/2025
Evaluador DPO/Jurídico	Oficial de Protección de Datos (DPO) / Área Jurídica	**Fecha Evaluación DPO/Jurídica**	26/11/2025
Representante Comité de IA	Representante del Comité de IA	**Fecha Revisión Final**	26/11/2025

Legalización y Firmas

Certificamos que esta evaluación se realizó conforme al **Checklist estandar** y que el dictamen final es el reflejo de la documentación y evidencia presentada por el proveedor **VisionTech OCR Services**.

Firma del Responsable Técnico

Firma del Representante del Comité de IA

Anexo E6. Formato de Model Card Diligenciado

Acta de Aprobación de Ficha Técnica del Modelo (Model Card)

RESIDENCIA_BOG_OCR_VALIDATOR_V1.0 (1.0 (Release Candidate)) | Entidad: Secretaría de Gobierno del Distrito Capital | Estado: PRUEBAS

Cargar Ficha Técnica del Modelo JSON

Seleccionar archivo | Ningún archivo seleccionado

Cargue el archivo JSON del Model Card para generar el acta.

Ficha Técnica del Modelo cargada para impresión. Puede imprimir ahora.

1. Información General

Nombre del modelo	RESIDENCIA_BOG_OCR_VALIDATOR_V1.0
Versión	1.0 (Release Candidate)
Entidad responsable	Secretaría de Gobierno del Distrito Capital
Área usuaria	Dirección de Atención al Ciudadano
Fecha de creación	2025-11-26
Fecha de actualización	2025-11-26
Estado del ciclo de vida	En pruebas

2. Propósito del Modelo

Objetivo del modelo	Extraer y validar automáticamente el nombre del solicitante, número de cédula y dirección del predio a partir de documentos en formato PDF o imagen (Cédula y Recibo de Servicio Público).
Caso de uso	Trámite administrativo de expedición del Certificado de Residencia en Bogotá.
Alcance	El modelo recomienda la validación de documentos. No rechaza automáticamente; los casos con confianza <90% se derivan a revisión humana.
Procesos institucionales impactados	Proceso de expedición del Certificado de Residencia, transformando la validación manual en una gestión por excepción.
Usuarios previstos	Ciudadanos solicitantes del certificado y funcionarios que gestionan la 'Bandeja de Excepciones'.

3. Descripción Técnica

Tipo de modelo	Visión por Computadora
Algoritmo principal	OCR (Reconocimiento Óptico de Caracteres) y NLP (Procesamiento de Lenguaje Natural).
Arquitectura	Modelo pre-entrenado de OCR ajustado (fine-tuning) con una taxonomía de recibos de servicios públicos de Bogotá (Enel, Acueducto, Vanti).
Tecnologías utilizadas	Python, TensorFlow/PyTorch, OpenCV.

4. Datos Utilizados

Dataset de Entrenamiento	80% del dataset RESIDENCIA_BOG_VALIDATION_V1.
Dataset de Validación	10% del dataset RESIDENCIA_BOG_VALIDATION_V1.
Dataset de Pruebas	10% del dataset RESIDENCIA_BOG_VALIDATION_V1.
Referencia al Data Sheet	Data Sheet RESIDENCIA_BOG_VALIDATION_V1 (versión 1.0).

5. Métricas de Desempeño

Métricas Globales

Métrica	Valor	Descripción
Precisión Global (Accuracy)	96.5%	Porcentaje de extracciones correctas de pares Nombre-Dirección en el conjunto de pruebas.
Tasa de Resolución Autónoma	>90% (Meta)	Porcentaje de casos que el sistema puede validar con una confianza >90% sin intervención humana.

Métricas por Subpoblaciones

Subpoblación	Métrica	Valor	Observaciones
Documentos Digitales / Alta Calidad	Precisión	98.5%	Grupo de referencia con el mejor rendimiento.
Fotos de Celular / Calidad Media	Precisión	95.0%	Rendimiento dentro de los márgenes aceptables.
Fotos de Baja Calidad / Borrosas	Precisión	82.0%	Rendimiento bajo, pero el riesgo se mitiga al activar el protocolo de desvío a humano para estos casos.

6. Evaluación de Sesgos

Sesgos identificados	Sesgo de representación hacia facturas digitales de Enel, corregido en el Data Sheet. Riesgo de sesgo técnico por calidad de imagen.
Resultados de pruebas de equidad	La diferencia en la tasa de error entre documentos de alta calidad (98.5%) y baja calidad (94.7%) fue del 3.8%, cumpliendo el umbral <5%.
Medidas aplicadas para mitigar sesgos	Balanceo del dataset de entrenamiento (oversampling de imágenes de baja calidad y recibos físicos) y protocolo 'Human-in-the-loop' para casos de baja confianza.

7. Riesgos del Modelo

Tipo de Riesgo	Descripción	Impacto	Probabilidad
Técnico	Posibilidad de 'Concept Drift' si los proveedores de servicios públicos cambian el formato de sus facturas.	Medio	Media
Ético	Riesgo de falsos negativos por condiciones de iluminación extremas (reflejo de flash), afectando la experiencia del usuario.	Bajo	Media

8. Controles Implementados

Tipo de Control	Descripción	Frecuencia	Responsable
Supervisión humana	Protocolo 'Human-in-the-loop' para todos los casos con confianza <90% o clasificados como 'No Válidos'.	Tiempo real	Funcionarios de Atención al Ciudadano
Monitoreo continuo	Dashboard de Gobernanza para monitorear KPIs de equidad y 'drift'.	Tiempo real / Revisión trimestral	Líder Técnico / Comité de IA

9. Explicabilidad y Transparencia

Técnicas de explicabilidad utilizadas	No se utilizan técnicas XAI complejas (LIME/SHAP). La explicabilidad se basa en la confianza del OCR y la derivación a humanos.
Información pública disponible	Se publicará una versión simplificada de esta Model Card en el portal de la entidad.
Mecanismos de rendición de cuentas	Auditorías trimestrales por parte del Comité de IA y un canal PQR específico para decisiones algorítmicas.

10. Reglas de Uso Responsable

Usos permitidos	Exclusivamente para la validación administrativa de residencia en el trámite distrital.
Usos restringidos	No aplica. Los usos no permitidos están prohibidos.
Usos prohibidos	Utilizar los datos extraídos (consumo, estrato) para evaluar capacidad de pago, realizar scoring crediticio o crear perfiles de comportamiento.
Buenas prácticas operativas	Privacidad por diseño (eliminación de datos post-trámite) y transparencia activa (aviso al ciudadano).

11. Entradas y Salidas del Modelo

Tipos de datos de entrada	Imágenes en formato PDF, JPG o PNG de la Cédula de Ciudadanía y un recibo de servicio público.
Tipos de predicciones/salidas generadas	JSON con los datos extraídos (nombre, ID, dirección) y un puntaje de confianza (0-1).

12. Monitoreo y Mantenimiento

Indicadores de desempeño	Precisión Global, Tasa de Resolución Autónoma, Diferencia de Tasa de Error (Equidad).
Frecuencia de reentrenamiento	Según necesidad, activado por monitoreo de 'Concept Drift' (ej. si un proveedor cambia el formato de su factura).
Controles de drift	Monitoreo de la tasa de derivación a humanos. Un aumento sostenido indica un posible 'drift' y activa una alerta para revisión.

13. Historial de Cambios

Versión	Fecha	Ajustes realizados	Responsable
1.0	2025-11-26	Versión inicial. Modelo ajustado (fine-tuned) con dataset balanceado.	Equipo de Desarrollo e Innovación

13. Historial de Cambios

Versión	Fecha	Ajustes realizados	Responsable
1.0	2025-11-26	Versión inicial. Modelo ajustado (fine-tuned) con dataset balanceado.	Equipo de Desarrollo e Innovación

Aprobación de Gobernanza y Despliegue

Aprobación Técnica	Aprobado. El modelo cumple con los umbrales de precisión (>95%) y es técnicamente robusto.
Aprobación Jurídica	Aprobado. El protocolo de intervención humana preserva el debido proceso.
Aprobación DPO/Privacidad	Aprobado. Los controles de privacidad por diseño y la prohibición de usos secundarios son adecuados.
Decisión Final del Comité de IA	Aprobado para despliegue (Go-Live), condicionado a monitoreo intensivo durante las primeras 4 semanas.

Legalización y Firmas

Certificamos que esta Ficha Técnica ha sido revisada y que el modelo de IA cumple con los requisitos de desempeño, equidad, privacidad y gobernanza para su estado actual (Producción/Pruebas).

Nombre Completo del Responsable Técnico del Modelo
Mario Alexander Ortiz Firma

Nombre del Representante del Comité de IA
Javier Mauricio Rocha Firma

ANEXO F. Manuales Funcionales

Anexo F1. Manual Funcional de AI Use-Case Canvas

Introducción:

Este anexo presenta el Manual Funcional del AI Use-Case Canvas, herramienta inicial del Framework de Gobernanza de Inteligencia Artificial del Distrito Capital. El documento está diseñado para ser utilizado por funcionarios públicos y ciudadanía en general, con lenguaje claro, enfoque práctico y alineado con los criterios académicos del Trabajo Final de Máster (UNIR).

Paso 1. Identificación del Caso de Uso

Objetivo del paso

Registrar la información básica que permite identificar el caso de uso, su responsable institucional y su trazabilidad dentro del Distrito Capital.

Campos de la herramienta y cómo diligenciarlos

- **Nombre del caso de uso**

  Escriba un nombre claro, concreto y comprensible que describa el propósito del proyecto.

  Ejemplo: "Priorización de solicitudes ciudadanas para atención social".

- **Entidad distrital responsable**

  Indique la entidad que lidera el caso de uso.

  Ejemplo: Secretaría Distrital de Gobierno.

- **Dependencia o área responsable**

  Registre la dependencia que gestiona el proceso donde se aplicará la IA.

  Ejemplo: Dirección de Atención al Ciudadano.

- **Responsable institucional**

  Nombre y cargo del servidor público responsable del caso de uso.

- **Fecha de formulación**

  Fecha en la que se diligencia el IA Use-Case Canvas.

Recomendaciones prácticas

- Evite nombres genéricos como “Proyecto IA”.
- Este paso no define la tecnología, solo el contexto institucional.

Forma 1. Sección 1 – Identificación del Caso de Uso del IA Use-Case Canvas.

Paso 2. Contexto y Propósito

Objetivo del paso

Describir el problema público o necesidad institucional que motiva el uso de inteligencia artificial.

Campos de la herramienta y cómo diligenciarlos

- **Descripción del problema**

  Explique la situación actual que se desea mejorar, en lenguaje claro y sin tecnicismos.

- **Proceso o servicio afectado**

  Indique el trámite, servicio o proceso institucional donde se presenta el problema.

- **Propósito del uso de IA**

  Describa qué se espera lograr con la IA (mejorar tiempos, calidad, cobertura, etc.).

- **Resultados esperados**

  Señale los beneficios concretos para la entidad y la ciudadanía.

Recomendaciones prácticas

- Siempre partir del problema, no de la solución tecnológica.
- Sea específico y evite generalidades.

Forma 2. Sección 2 – Contexto y Propósito del IA Use-Case Canvas.

Paso 3. Actores Involucrados

Objetivo del paso

Identificar todas las personas, áreas y grupos que participan o se ven impactados por el caso de uso.

Campos de la herramienta y cómo diligenciarlos

- **Usuarios del sistema**

  Funcionarios que utilizarán directamente la solución.

- **Ciudadanía o grupos impactados**

  Personas o colectivos que se verán beneficiados o afectados.

- **Áreas responsables**

  Dependencias con responsabilidades en operación, control o supervisión.

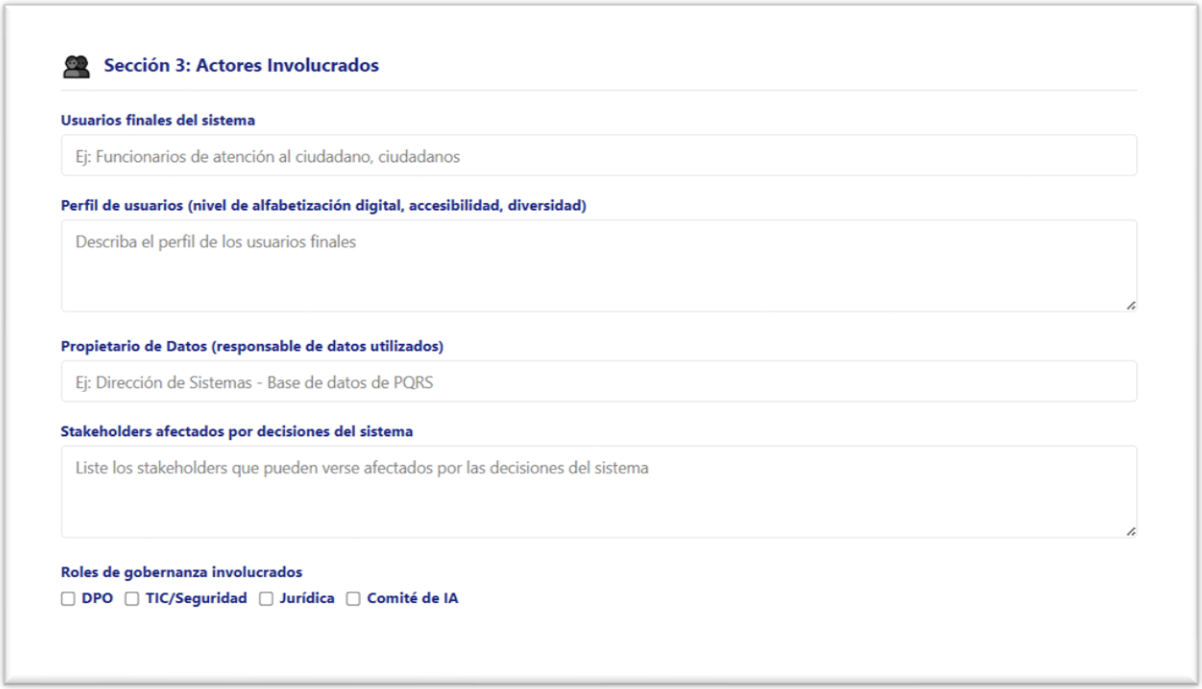
- **Otros actores relevantes**

  Entidades externas o aliados, si aplica.

Recomendaciones prácticas

- Diferencie claramente entre operadores y beneficiarios.
- Considere impactos directos e indirectos.

Forma 3. Sección 3 – Actores Involucrados del IA Use-Case Canvas.



Paso 4. Datos Requeridos

Objetivo del paso

Identificar los datos necesarios para el funcionamiento del caso de uso y evaluar su sensibilidad.

Campos de la herramienta y cómo diligenciarlos

- **Tipos de datos requeridos**

Indique si son datos personales, no personales o sensibles.

- **Fuentes de los datos**

  Señale si provienen de sistemas internos, bases externas o mixtas.

- **Volumen y periodicidad**

  Describa la cantidad aproximada y frecuencia de actualización.

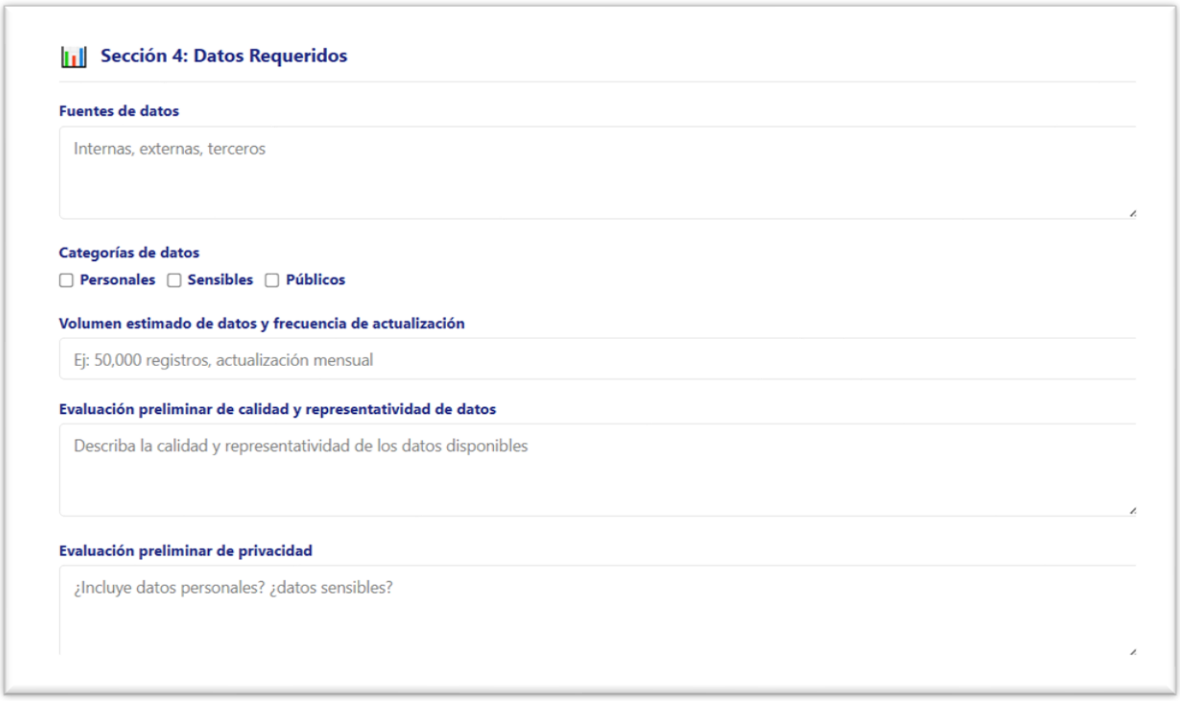
- **Consideraciones de calidad**

  Indique si los datos son completos, actualizados y confiables.

Recomendaciones prácticas

- Este paso no autoriza el uso de datos, solo los identifica.
- Sea transparente sobre limitaciones de calidad.

Forma 4. Sección 4 – Datos Requeridos del IA Use-Case Canvas.



Paso 5. Revisión Legal

Objetivo del paso

Verificar la viabilidad jurídica del caso de uso y su alineación con la normativa vigente.

Campos de la herramienta y cómo diligenciarlos

- **Base legal del servicio o proceso**

  Cite la norma que habilita el servicio público.

- **Base legal para el tratamiento de datos**

  Indique si aplica Ley 1581 de 2012 u otra norma relevante.

- **Restricciones legales identificadas**

  Señale posibles límites o requerimientos adicionales.

Recomendaciones prácticas

- No se requiere análisis jurídico profundo, solo identificación preliminar.
- Este paso previene riesgos legales posteriores.

Forma 5. Sección 5 – Revisión Legal del IA Use-Case Canvas.

Paso 6. Clasificación de Riesgo

Objetivo del paso

Clasificar el nivel de riesgo del sistema de IA según su impacto potencial.

Campos de la herramienta y cómo diligenciarlos

- **Nivel de riesgo**

  Seleccione: mínimo, limitado o alto.

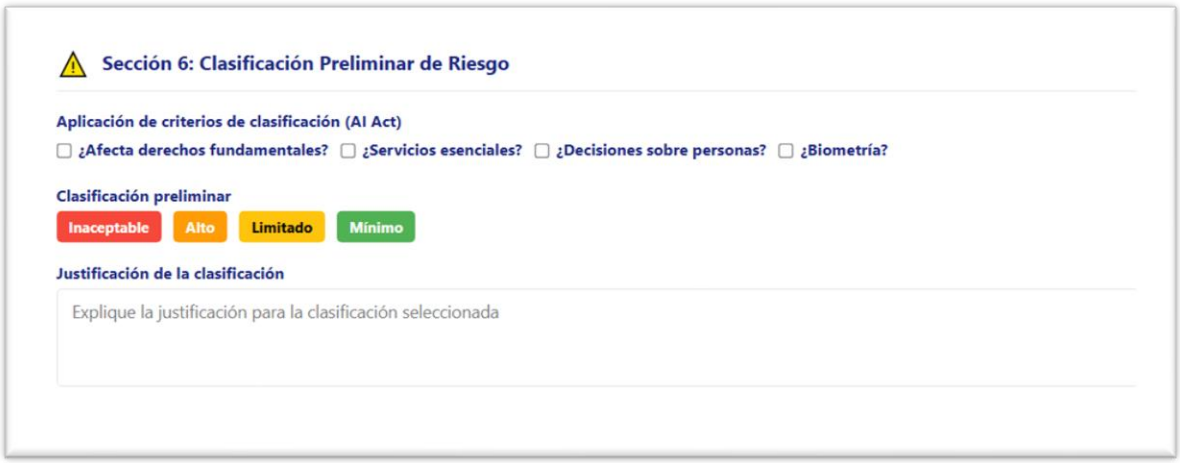
- **Justificación de la clasificación**

  Explique brevemente por qué se asigna ese nivel.

Recomendaciones prácticas

- Ante la duda, clasifique de manera conservadora.
- Esta clasificación define exigencias de gobernanza posteriores.

Forma 6. Sección 6 – Clasificación de Riesgo del IA Use-Case Canvas.



Paso 7. Identificación de Riesgos

Objetivo del paso

Reconocer riesgos éticos, legales, técnicos y operativos asociados al caso de uso.

Campos de la herramienta y cómo diligenciarlos

- **Riesgos éticos**

  Describa posibles riesgos relacionados con sesgos, discriminación, trato desigual o afectación a derechos fundamentales.

  Ejemplo: riesgo de priorizar de forma injusta solicitudes de ciertos grupos poblacionales.

- **Riesgos de privacidad y protección de datos**

  Indique posibles riesgos asociados al uso, tratamiento o almacenamiento de datos personales.

  Ejemplo: uso de datos personales sin la debida minimización.

- **Riesgos técnicos u operativos**

  Identifique riesgos relacionados con fallas del sistema, errores de funcionamiento o dependencia tecnológica.

Ejemplo: interrupciones del servicio por fallos del modelo.

- **Riesgos reputacionales o institucionales**

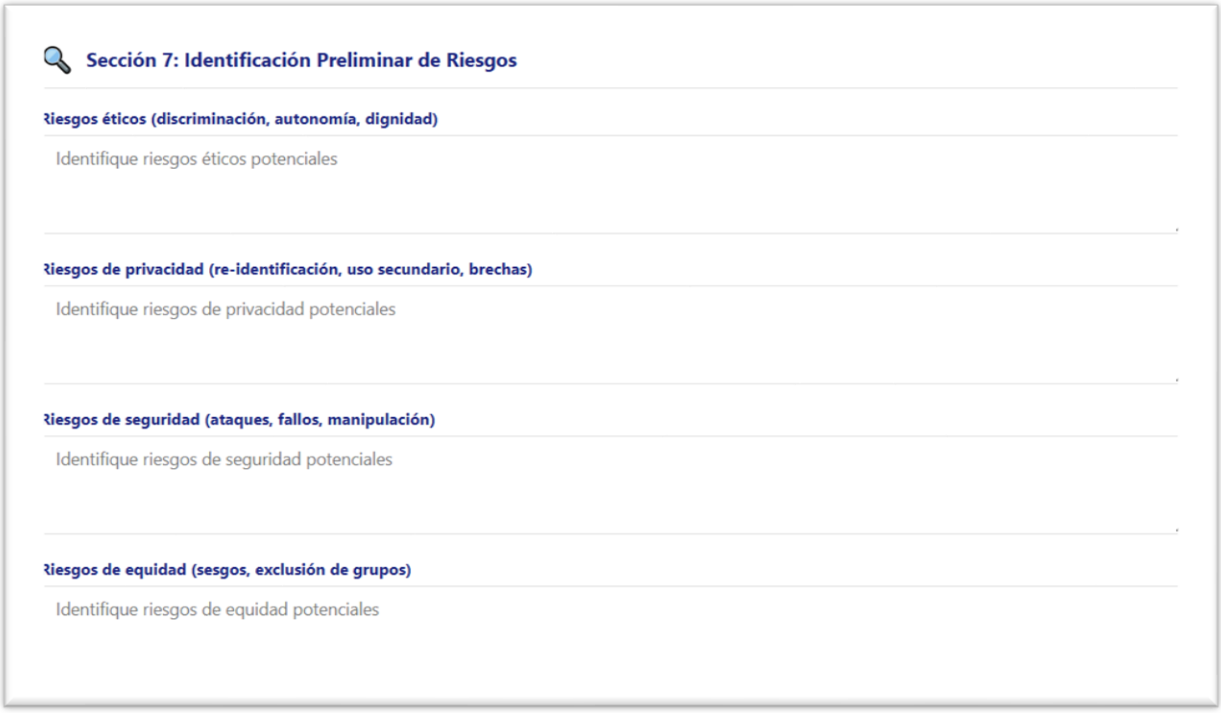
  Describa riesgos que puedan afectar la confianza ciudadana o la imagen de la entidad.

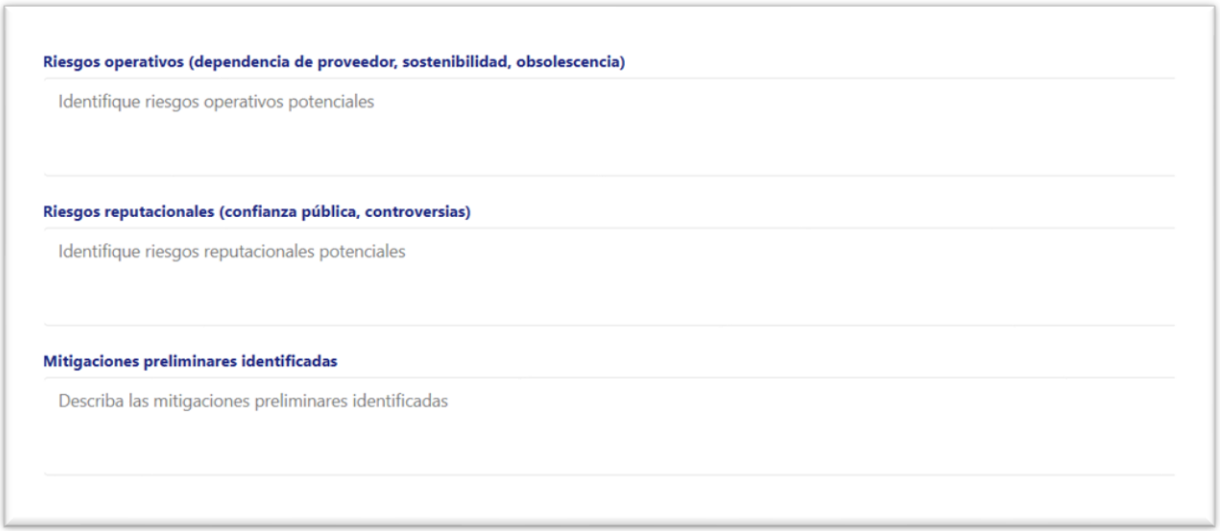
  Ejemplo: percepción de decisiones automatizadas injustas.

Recomendaciones prácticas

- No se espera eliminar riesgos, sino identificarlos.
- La transparencia es clave para la gobernanza.

Forma 7. Sección 7 – Identificación de Riesgos del IA Use-Case Canvas.





Paso 8. Métricas de Éxito

Objetivo del paso

Definir cómo se evaluará si el caso de uso genera valor público.

Campos de la herramienta y cómo diligenciarlos

- **Indicadores de desempeño del sistema**

  Defina métricas que midan el funcionamiento general del caso de uso.

  Ejemplo: reducción de tiempos de respuesta.

- **Indicadores de impacto en el servicio público**

  Señale métricas relacionadas con la mejora del servicio a la ciudadanía.

  Ejemplo: aumento en la satisfacción ciudadana.

- **Indicadores de cumplimiento ético y normativo**

  Incluya métricas que permitan verificar el cumplimiento de principios de equidad, transparencia y legalidad.

  Ejemplo: ausencia de reclamaciones por sesgos.

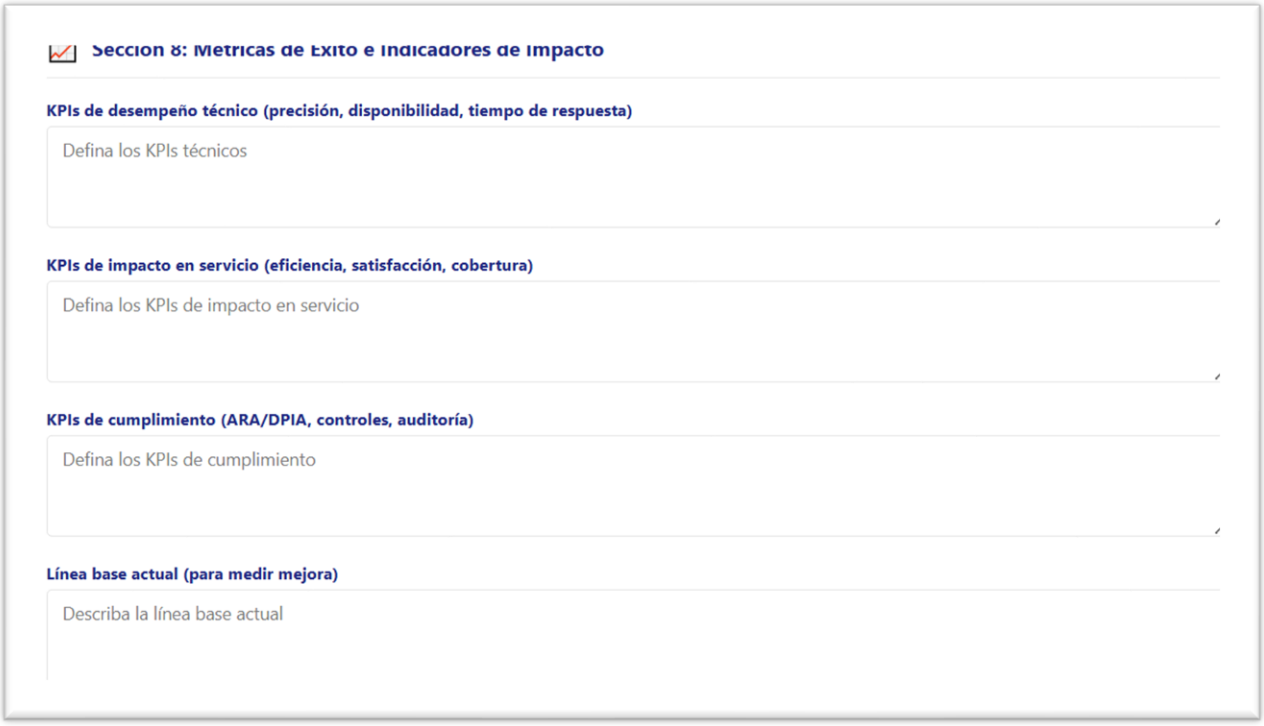
- **Línea base y meta esperada**

  Indique el estado actual y la meta a alcanzar.

Recomendaciones prácticas

- Priorice métricas simples y comprensibles.
- Evite indicadores exclusivamente técnicos.

Forma 8. Sección 8 – Métricas de Éxito del IA Use-Case Canvas.



Paso 9. Plan de Despliegue

Objetivo del paso

Describir cómo se implementará el caso de uso de forma gradual y controlada.

Campos de la herramienta y cómo diligenciarlos

- **Estrategia de implementación**

  Describa cómo se pondrá en marcha el sistema (piloto, prueba controlada, despliegue progresivo).

- **Fases del despliegue**

  Indique las etapas previstas y su duración aproximada.

- **Capacitación de usuarios**

  Señale cómo se capacitará a funcionarios u otros usuarios del sistema.

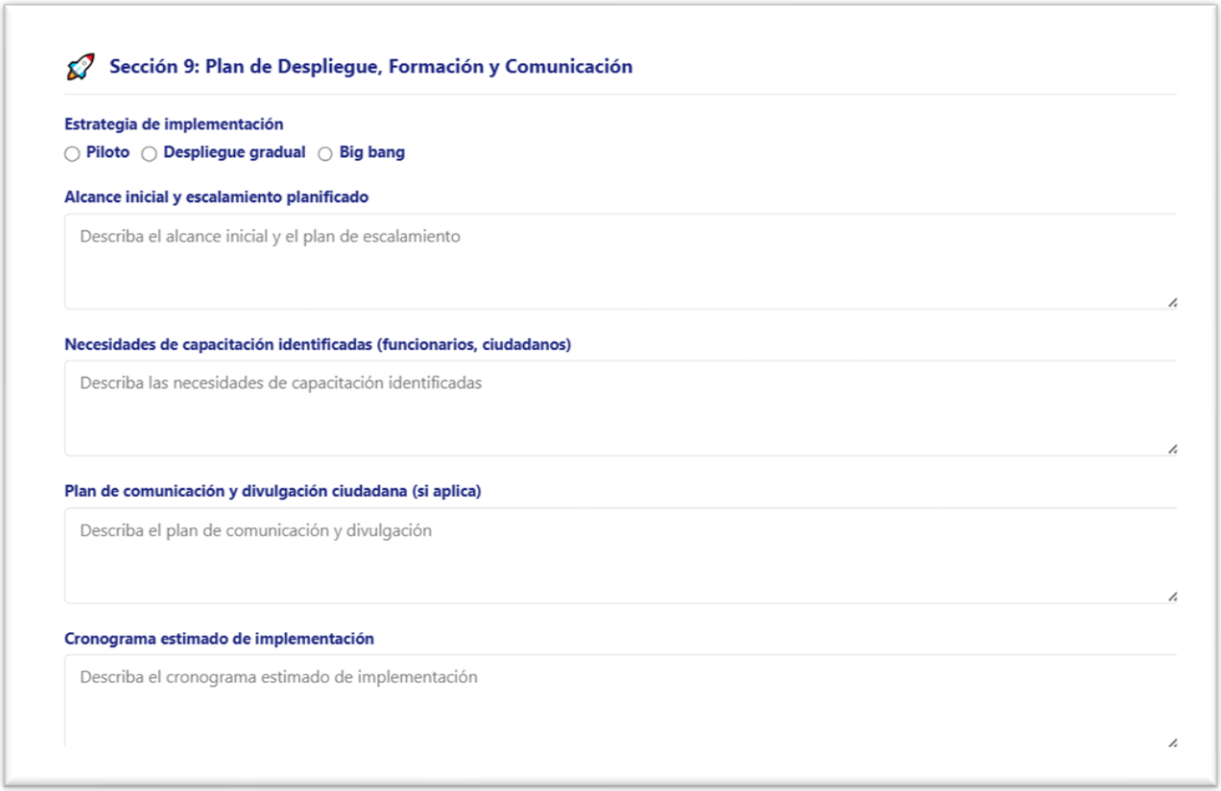
- **Estrategia de comunicación**

  Describa cómo se informará a la ciudadanía o a los usuarios internos sobre el uso del sistema.

Recomendaciones prácticas

- Priorice pilotos antes de despliegues masivos.
- Incluya gestión del cambio organizacional.

Forma 9. Sección 9 – Plan de Despliegue del IA Use-Case Canvas.



Paso 10. Monitoreo y Auditoría

Objetivo del paso

Establecer mecanismos de seguimiento continuo del sistema.

Campos de la herramienta y cómo diligenciarlos

- **Responsables del monitoreo**

  Indique el área o dependencia encargada del seguimiento del sistema.

- **Frecuencia de revisión**

  Defina cada cuánto se evaluará el desempeño y los impactos del sistema.

- **Mecanismos de auditoría**

  Describa cómo se realizarán revisiones internas o externas.

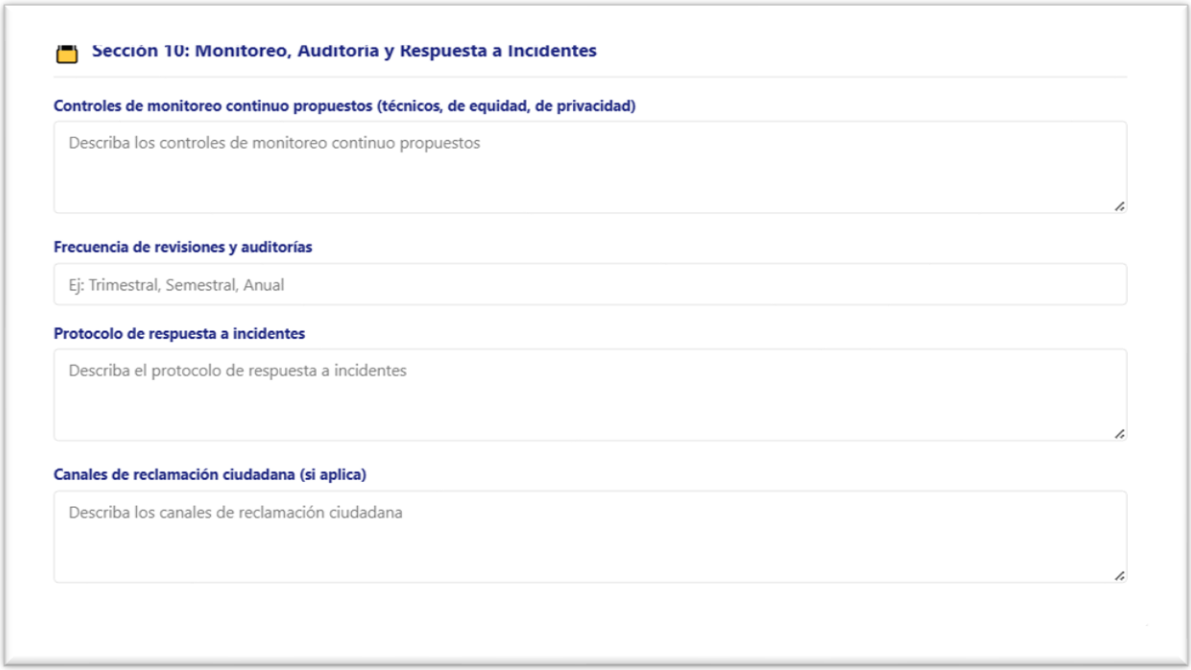
- **Canales de reporte y reclamación**

  Indique cómo la ciudadanía o los usuarios pueden reportar errores o inconformidades.

Recomendaciones prácticas

- El monitoreo debe ser permanente.
- Permite detectar impactos no previstos.

Forma 10. Sección 10 – Monitoreo y Auditoría del IA Use-Case Canvas.



Paso 11. Plan de Fin de Vida

Objetivo del paso

Definir cómo se gestionará la desactivación o sustitución del sistema de IA.

Campos de la herramienta y cómo diligenciarlos

- **Condiciones de finalización del sistema**

  Indique cuándo y por qué se dejaría de utilizar el sistema.

- **Gestión de datos al cierre**

  Describa qué ocurrirá con los datos una vez finalizado el uso (borrado, archivo, anonimización).

- **Continuidad del servicio**

  Explique cómo se garantizará el servicio sin el sistema de IA.

Recomendaciones prácticas

- Todo sistema debe tener un cierre planificado.
- Evita dependencias tecnológicas irreversibles.

Forma 11. Sección 11 – Plan de Fin de Vida del IA Use-Case Canvas.

Paso 12. Aprobaciones

Objetivo del paso

Formalizar la validación institucional del caso de uso.

Campos de la herramienta y cómo diligenciarlos

- **Aprobación técnica**

  Registro del visto bueno del área de tecnologías.

- **Aprobación jurídica y de datos**

  Validación por parte de las áreas jurídica y de protección de datos.

- **Aprobación de gobernanza / Comité de IA**

  Decisión final del órgano de gobernanza correspondiente.

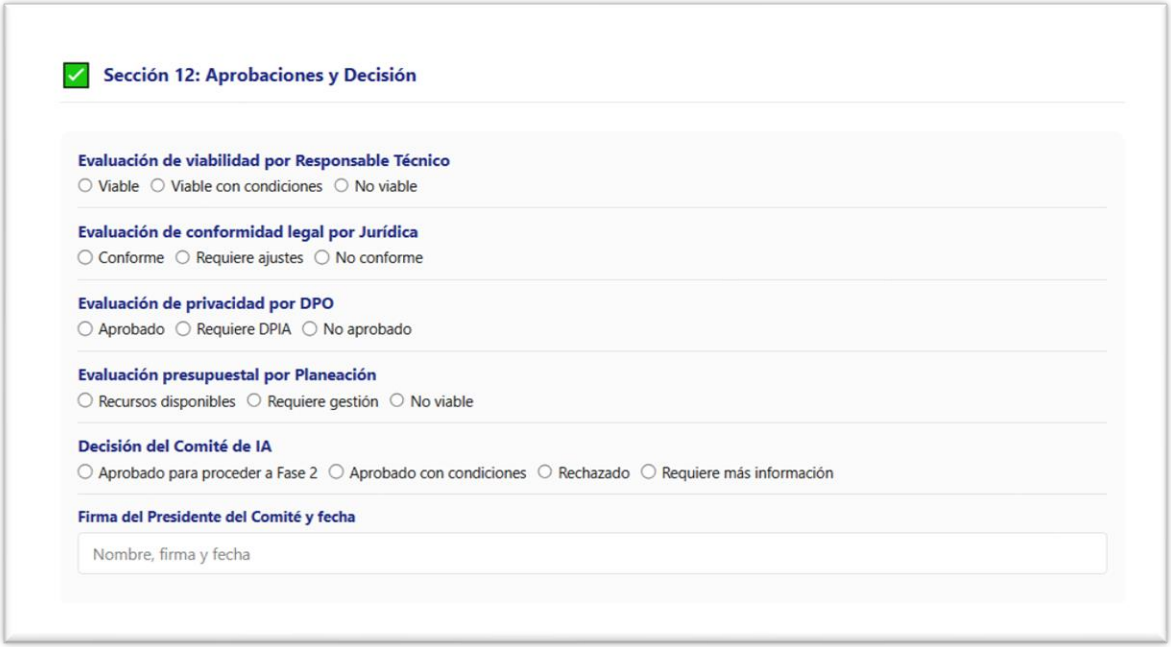
- **Observaciones finales**

  Comentarios o condiciones para continuar el proyecto.

Recomendaciones prácticas

- No avanzar sin aprobaciones mínimas.
- Este paso deja evidencia de debida diligencia.

Forma 12. Sección 12 – Aprobaciones del IA Use-Case Canvas.



Cierre del Manual

El diligenciamiento adecuado del IA Use-Case Canvas constituye un requisito esencial para la adopción responsable de inteligencia artificial en el Distrito Capital. Este manual debe utilizarse como referencia práctica para garantizar decisiones informadas, transparentes y alineadas con el interés público.

Anexo F2. Manual Funcional de la Matriz De Riesgos de IA

Guía oficial para diligenciar un nuevo riesgo en la herramienta del Framework de Gobernanza de IA del Distrito Capital.

Introducción:

La identificación temprana y sistemática de riesgos en proyectos de Inteligencia Artificial (IA) es un componente esencial de la gobernanza responsable en entidades públicas. En el contexto del Distrito Capital, donde los sistemas de IA pueden impactar derechos fundamentales, la asignación de beneficios públicos, la prestación de servicios esenciales o la toma de decisiones administrativas, resulta indispensable contar con un mecanismo estructurado que permita anticipar, evaluar y gestionar posibles afectaciones antes de que se materialicen.

Definir los riesgos no solo fortalece la transparencia y la rendición de cuentas, sino que también garantiza que la implementación de tecnologías de IA esté alineada con el marco normativo colombiano (Ley 1581 de 2012, Decreto 1377 de 2013, lineamientos de la SIC, CONPES 4144, Política Distrital Bogotá Territorio Inteligente, entre otros). Asimismo, facilita que las entidades adopten decisiones informadas respecto al uso de datos, seguridad, privacidad y equidad, reduciendo la probabilidad de sesgos algorítmicos, fallos operativos o impactos negativos en la ciudadanía.

La Matriz de Riesgos de IA es una herramienta práctica diseñada para acompañar a equipos técnicos, jurídicos y misionales en todas las fases del ciclo de vida de un sistema de IA. Su propósito es permitir que cualquier funcionario —independientemente de su perfil técnico— pueda documentar, priorizar y gestionar los riesgos de manera clara, uniforme y verificable. Con ello, no solo se protege a la ciudadanía, sino que también se fortalece la confianza institucional y la capacidad del Distrito para implementar soluciones de IA responsables, seguras y éticamente alineadas con el interés público.

1. Descripción del Riesgo

Explique brevemente en qué consiste el riesgo identificado. Use lenguaje claro y concreto. Indique qué puede salir mal y cómo afectaría la operación o a la ciudadanía.

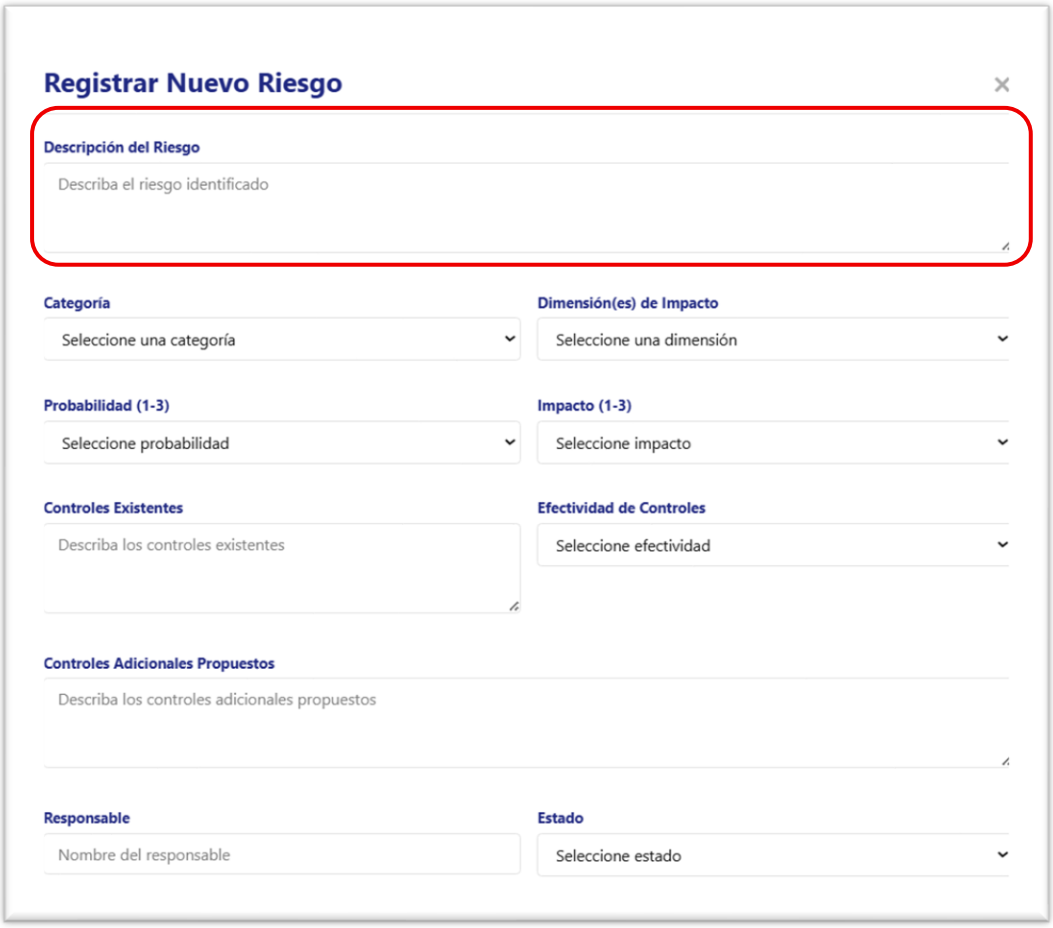
Cómo redactarlo

- Use lenguaje claro, sin tecnicismos.
- Explique qué podría salir mal y por qué es un riesgo.
- Evite frases vagas como “puede fallar”.

Ejemplos

- “El modelo puede generar recomendaciones sesgadas contra adultos mayores.”
- “Existe riesgo de fuga de datos sensibles durante el entrenamiento.”

Forma 13. Identificación Riesgo – Matriz de Riesgos de IA



2. Categoría del Riesgo

Debe seleccionar una de las categorías definidas por el framework:

- Privacidad
- Equidad
- Seguridad
- Transparencia
- Rendición de cuentas
- Continuidad del servicio
- Cumplimiento
- Reputación

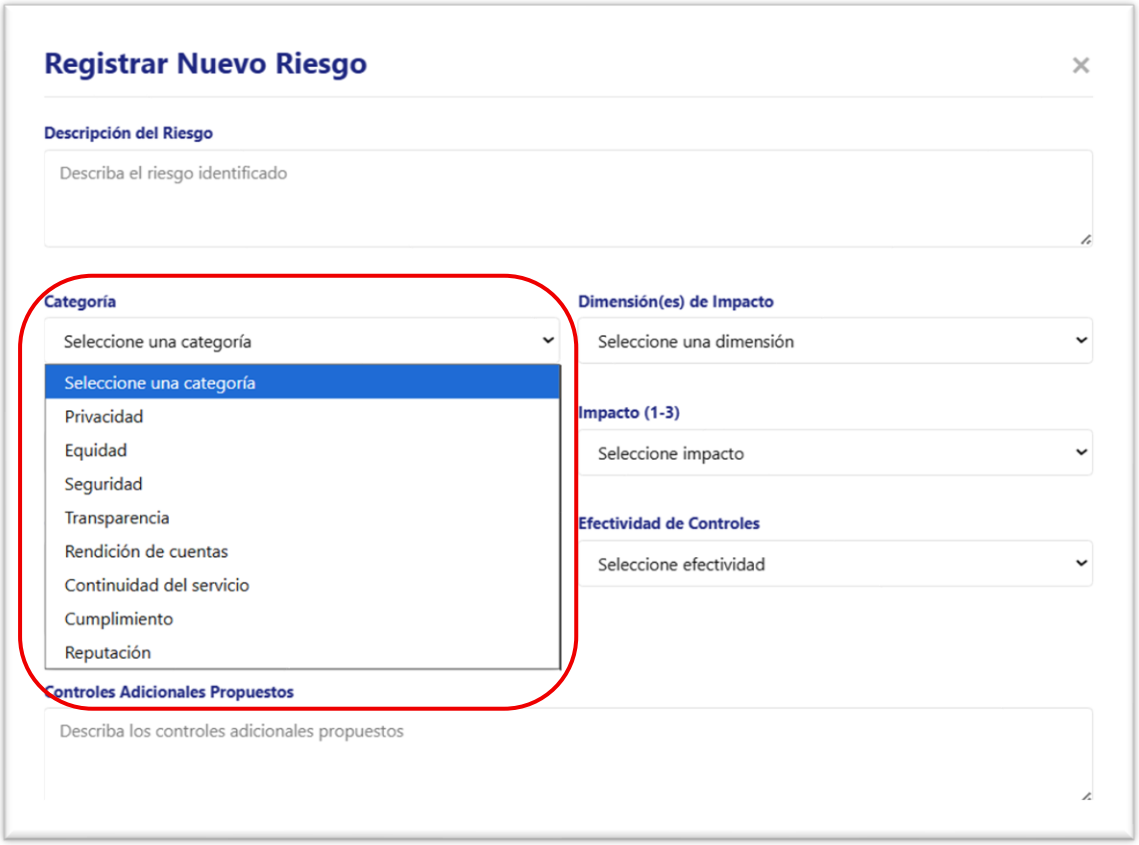
¿Cómo escogerla?

Seleccione la categoría que mejor represente la naturaleza del riesgo.

Recomendaciones

- Si afecta datos personales: Privacidad
- Si puede discriminar o excluir: Equidad
- Si compromete sistemas o infraestructura: Seguridad
- Si afecta claridad en decisiones: Transparencia
- Si implica trazabilidad o supervisión: Rendición de cuentas
- Si afecta disponibilidad del servicio: Continuidad
- Si tiene implicaciones legales: Cumplimiento
- Si puede dañar la confianza pública: Reputación

Forma 14. Categoría del Riesgo – Matriz de Riesgos de IA



3. Dimensiones del Impacto

Seleccione las dimensiones que pueden verse afectadas:

- Derechos y libertades fundamentales
- Equidad y no discriminación
- Continuidad y calidad del servicio
- Seguridad y ciberseguridad
- Reputación y confianza pública
- Cumplimiento regulatorio

Cómo diligenciarlo

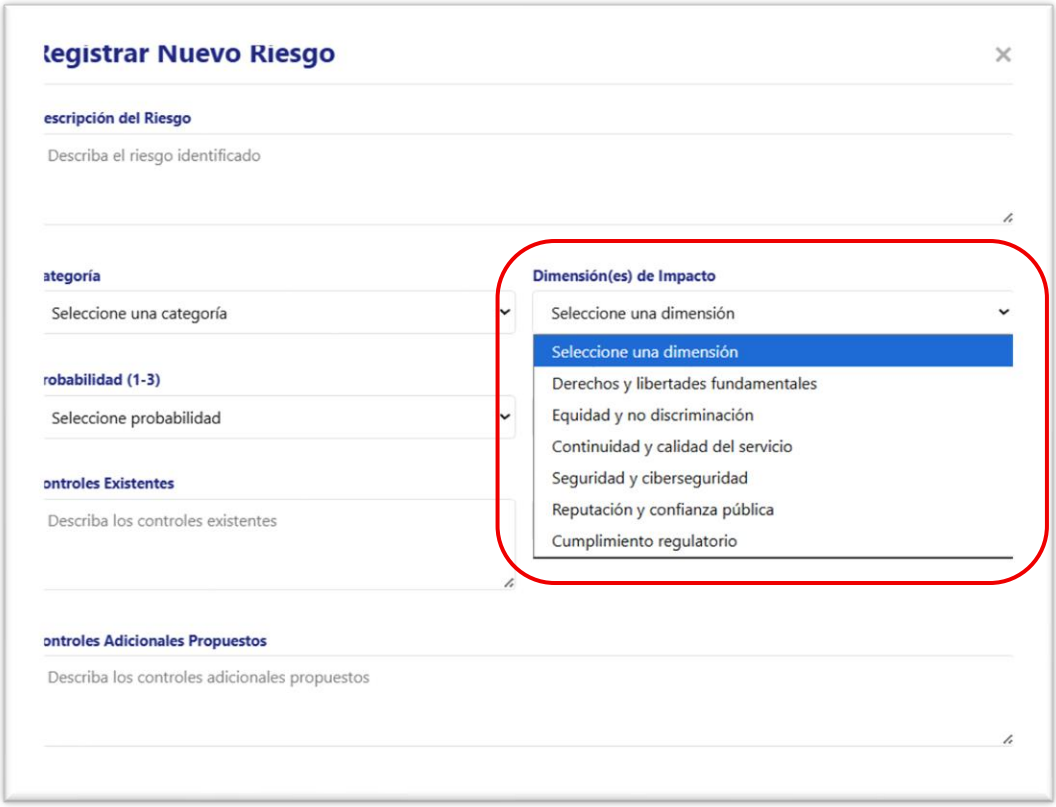
Elija una o varias dimensiones según corresponda.

Ejemplos

- Riesgo de sesgo → Equidad y no discriminación

- Riesgo de brecha de datos → Seguridad y ciberseguridad + Cumplimiento regulatorio
- Riesgo de mala clasificación de beneficiarios → Derechos fundamentales + Continuidad del servicio

Forma 15. Registro Nuevo Riesgo – Matriz de Riesgos de IA



4. Probabilidad

Se debe seleccionar:

- 1 = Baja
- 2 = Media
- 3 = Alta

Cómo decidir

Base la probabilidad en:

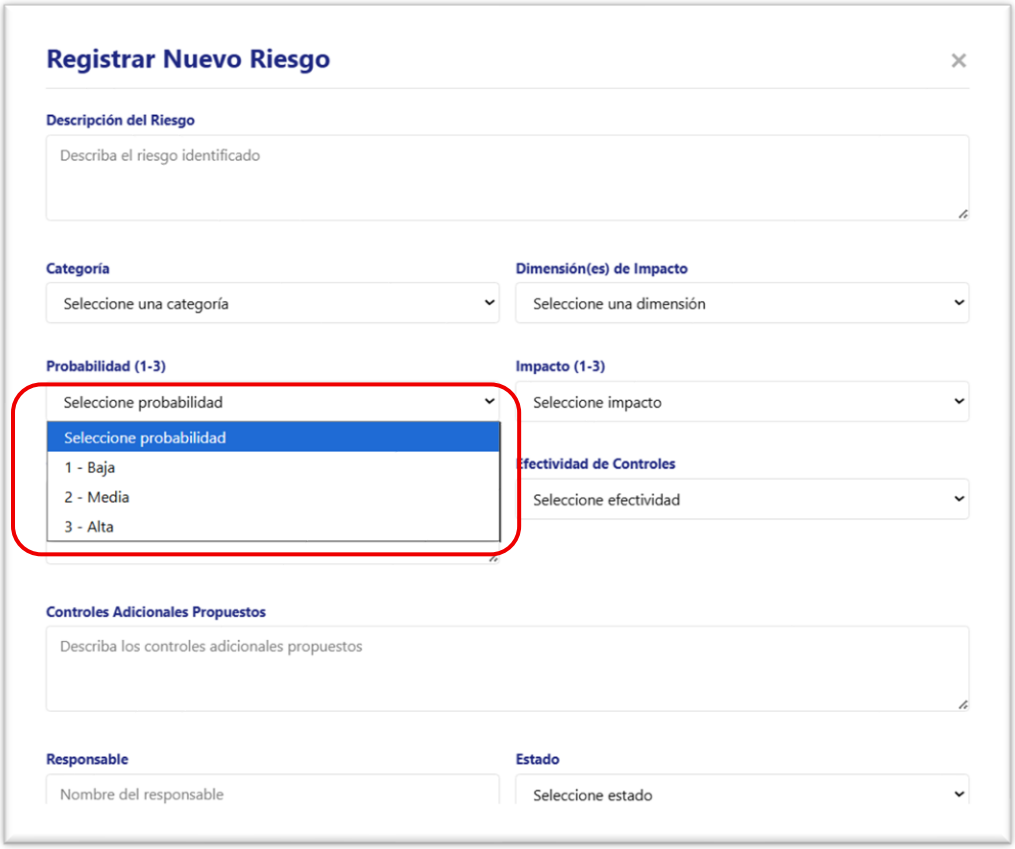
- Datos históricos
- Experiencia de la entidad

- Complejidad técnica del modelo
- Controles existentes

Ejemplos

- Modelo nuevo, sin pruebas suficientes → Probabilidad alta
- Riesgo ya mitigado por sistemas robustos → Probabilidad baja

Forma 16. Probabilidad de Riesgo – Matriz de Riesgos de IA



5. Impacto

Seleccione:

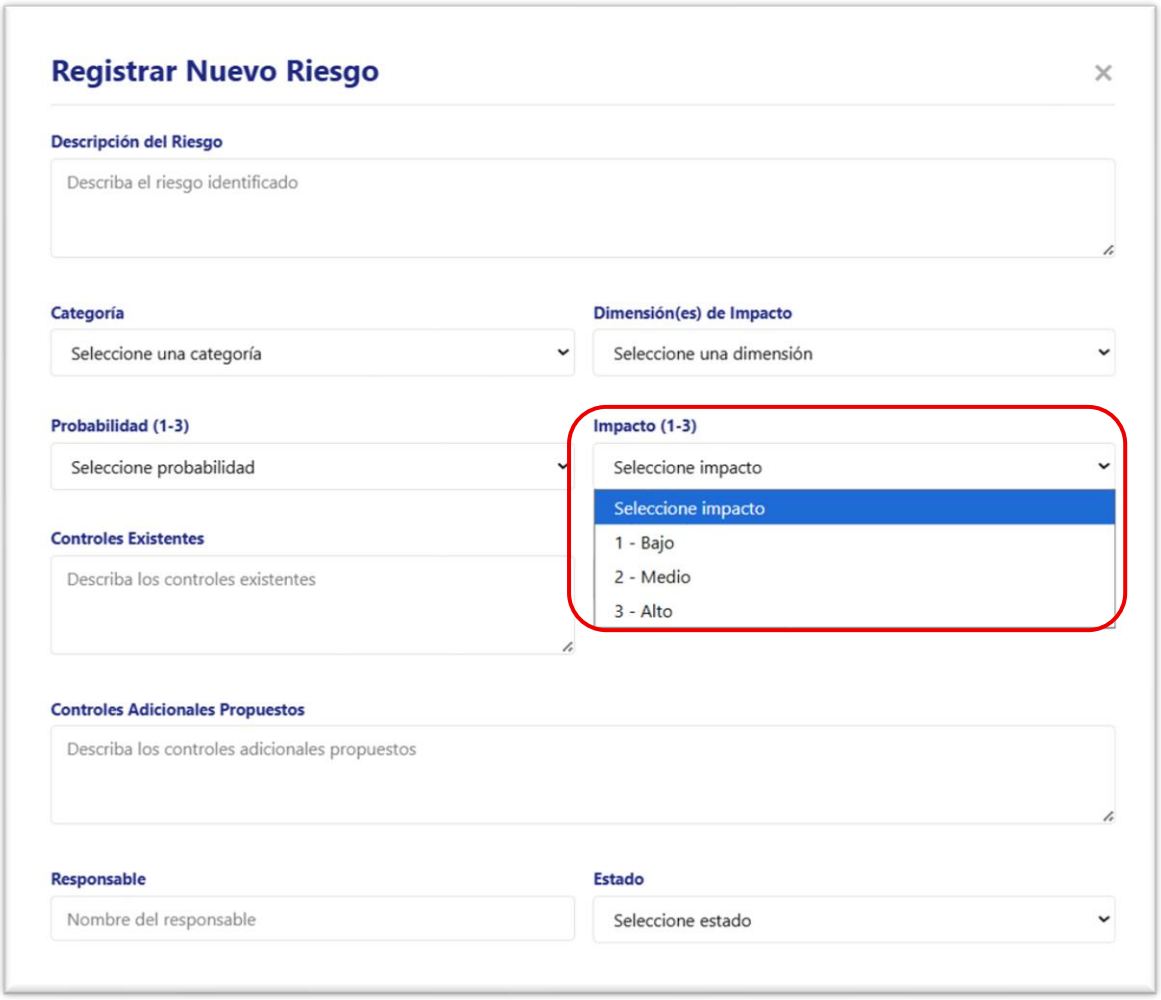
- **1 = Bajo**
- **2 = Medio**
- **3 = Alto**

Cómo decidir

Evalúe el efecto real si el riesgo se materializara:

- **Bajo:** afectación menor y manejable internamente.
- **Medio:** afecta calidad del servicio o genera reclamos.
- **Alto:** compromete derechos fundamentales, continuidad del servicio o confianza pública.

Forma 17. Impacto del Riesgo – Matriz de Riesgos de IA



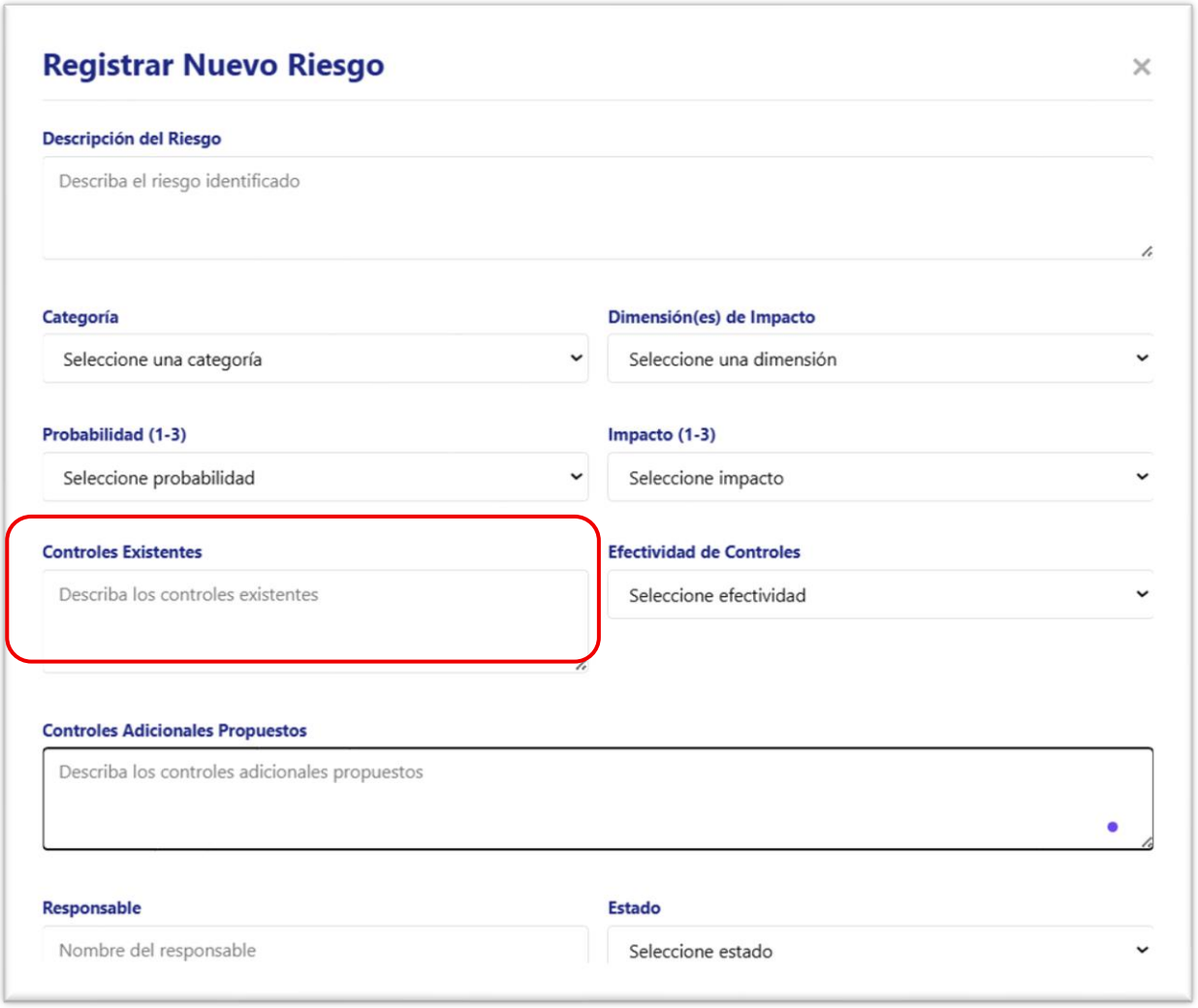
6. Controles Existentes

Describa las medidas que actualmente reducen o mitigan el riesgo.

Ejemplos

- “Validación humana en todas las decisiones críticas.”
- “Sistema con cifrado en tránsito y en reposo.”
- “Muestreo estadístico semestral de sesgos.”

Forma 18. Controles Existentes – Matriz de Riesgos de IA



7. Efectividad de Controles

Seleccione:

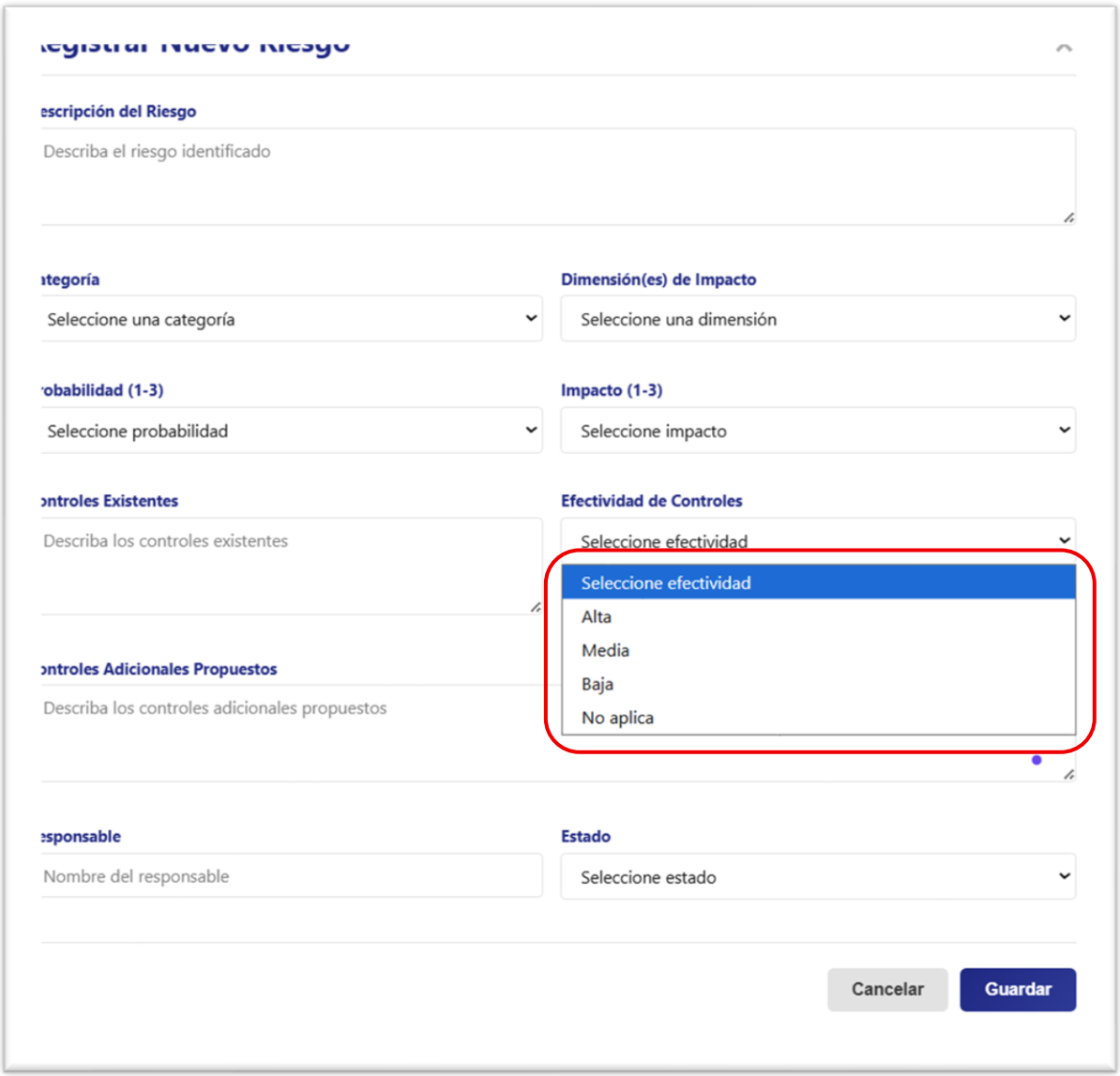
- **Alta**
- **Media**
- **Baja**
- **No aplica**

Cómo decidir

- **Alta:** el control reduce el riesgo de forma significativa.
- **Media:** ayuda, pero tiene limitaciones.

- **Baja:** tiene poca capacidad de mitigación.
- **No aplica:** no hay controles o no son relevantes.

Forma 19. Registro Nuevo Riesgo – Matriz de Riesgos de IA



8. Controles Adicionales Propuestos

Medidas nuevas o mejoras necesarias para reducir el riesgo.

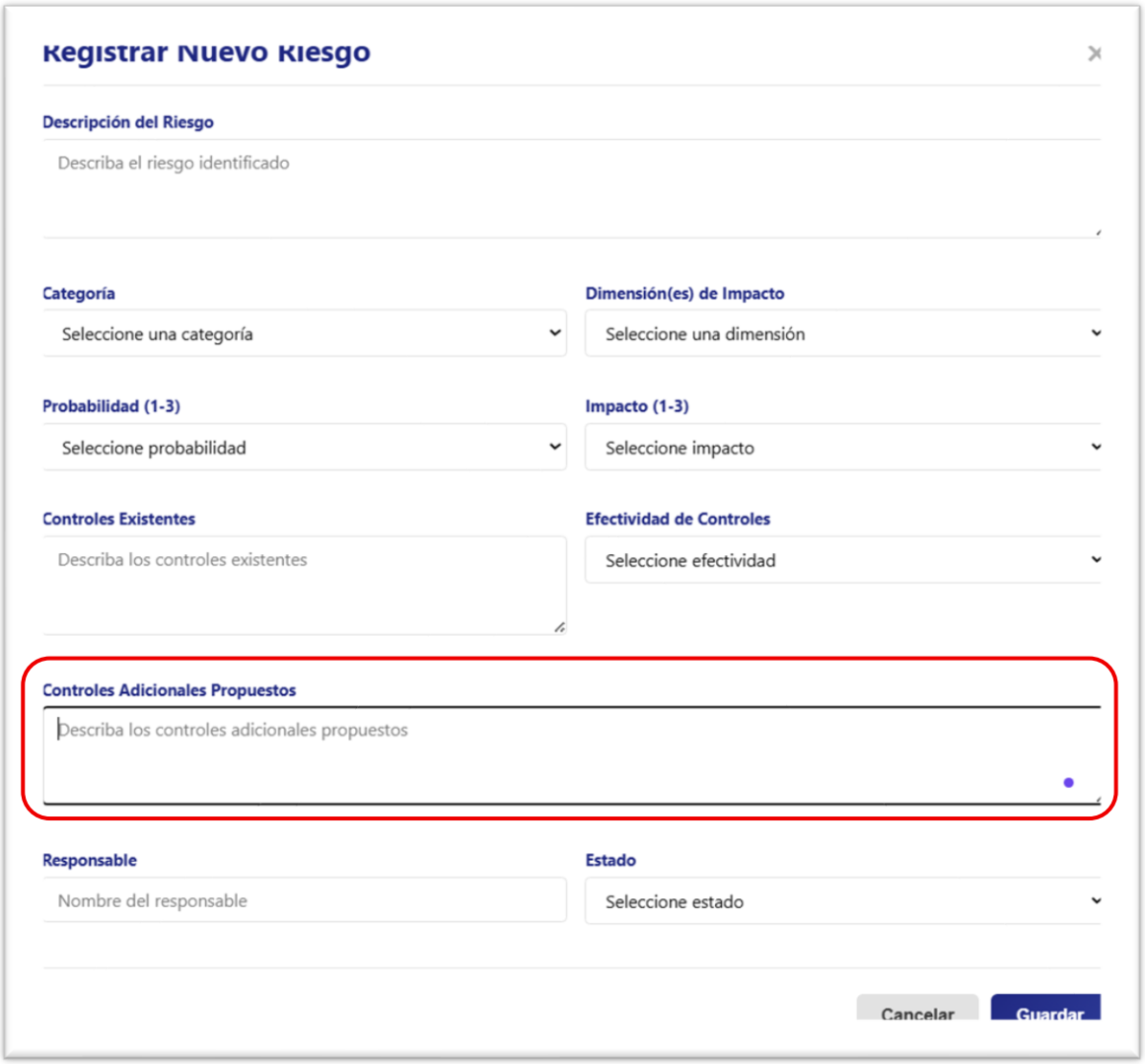
Ejemplos

- Implementar auditoría de equidad trimestral.
- Migración a infraestructura segura del Distrito.
- Capacitación de operadores en supervisión humana.

Consejo

Sea concreto: cada control debe poder ejecutarse y verificarse.

Forma 20. Controles Adicionales Propuestos – Matriz de Riesgos de IA



9. Nombre del Responsable

La persona o área que se encargará de gestionar este riesgo.

Opciones típicas

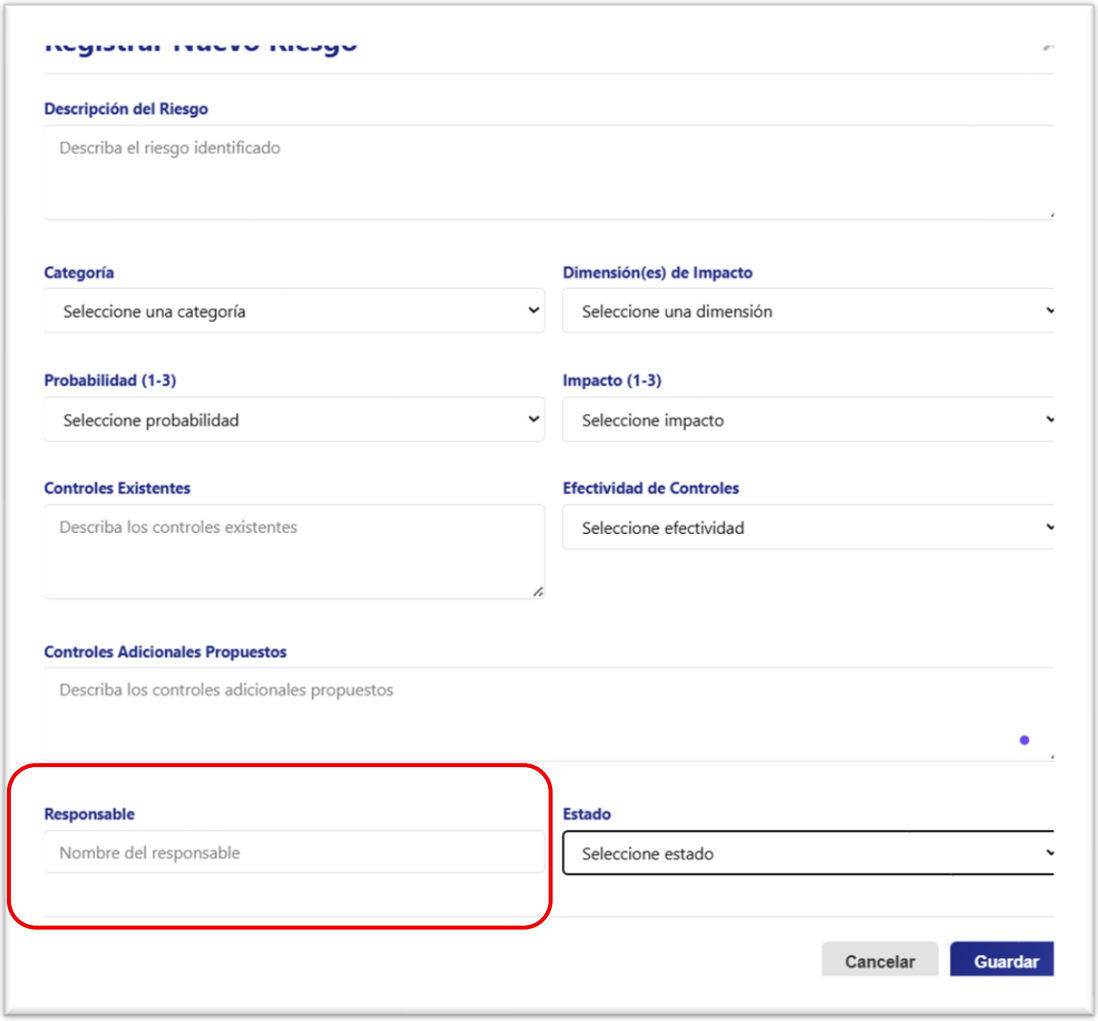
- Equipo de Analítica
- Dirección de Tecnología
- Oficina de Datos
- Oficina Jurídica

- Dependencia Operativa

Recomendación

Debe ser un responsable real capaz de liderar el plan de acción.

Forma 21. Nombre del Responsable – Matriz de Riesgos de IA



10. Estado del Riesgo

Debe seleccionar una de estas opciones:

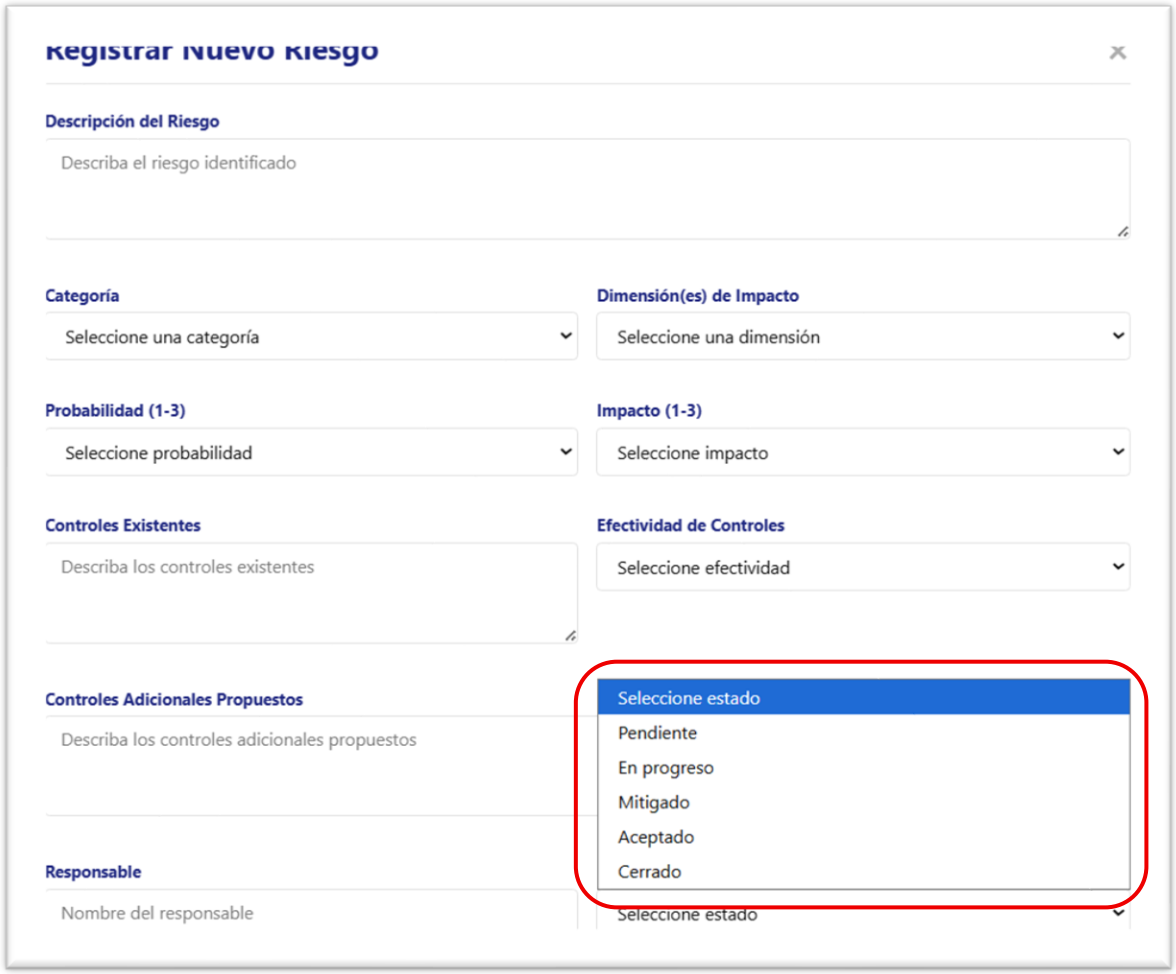
- Pendiente
- En progreso
- Mitigado
- Aceptado

- Cerrado

Cómo decidir

- Pendiente: recién identificado, sin acciones iniciadas.
- En progreso: hay acciones de mitigación en curso.
- Mitigado: se aplicaron controles y el riesgo se redujo a un nivel aceptable.
- Aceptado: el riesgo se reconoce, pero se acepta (cuando la mitigación es inviable).
- Cerrado: ya no existe o dejó de ser relevante.

Forma 22. Estado del Riesgo – Matriz de Riesgos de IA



Anexo F3. Manual Funcional del Formulario ARA/DPIA

Guía oficial para diligenciar un nuevo riesgo en la herramienta del Framework de Gobernanza de IA del Distrito Capital.

Introducción

La herramienta ARA/DPIA (Algoritmic Risk Assessment / Data Protection Impact Assessment) es un instrumento oficial del Framework Distrital de Gobernanza de Inteligencia Artificial que permite identificar, evaluar y controlar los riesgos asociados al uso de sistemas de IA en entidades públicas del Distrito Capital. Su propósito es asegurar que toda iniciativa basada en IA —desde prototipos hasta sistemas en producción— se desarrolle con altos estándares de ética, transparencia, protección de datos personales y respeto por los derechos fundamentales de la ciudadanía.

En la práctica, la herramienta funciona como un formulario guiado que orienta paso a paso a los equipos técnicos, jurídicos y misionales para entender el sistema de IA, los datos que utiliza, sus impactos y los riesgos que podría generar. La herramienta facilita que las entidades distritales puedan anticipar y mitigar riesgos como discriminación algorítmica, vulneración de la privacidad, fallas de seguridad, decisiones automatizadas sin supervisión humana adecuada o impactos desproporcionados en poblaciones vulnerables.

El formulario ARA/DPIA está inspirado en buenas prácticas internacionales —como el *AI Risk Management Framework* (NIST), el *AI Act* europeo y guías de Data Protection Impact Assessment—, pero adaptado rigurosamente al marco normativo colombiano, incluyendo la Ley 1581 de 2012, el Decreto 1377 de 2013, la Ley 1712 de 2014, la Ley 1437 de 2011, así como las directrices del CONPES 4144 de Política Nacional de IA y el CONPES Distrital 29 de Bogotá Territorio Inteligente.

La herramienta es especialmente útil porque:

- Estandariza la evaluación de riesgos en todas las entidades del Distrito.

- Facilita decisiones informadas antes de aprobar un proyecto de IA.
- Articula jurídica, técnica y operacionalmente a todas las dependencias involucradas.
- Asegura trazabilidad, dejando evidencia documentada de que se evaluaron riesgos antes de implementar tecnología.
- Promueve transparencia y confianza pública, al exigir mecanismos de explicabilidad y comunicación a la ciudadanía.
- Fortalece la gobernanza institucional, al exigir supervisión humana, auditoría periódica y medidas de mitigación.

Finalmente, el ARA/DPIA se convierte en una pieza central del ciclo de vida del framework, porque determina si un proyecto de IA puede avanzar, si requiere ajustes o si no debe implementarse. Por ello, su diligenciamiento es obligatorio para cualquier sistema de IA que se utilice en el Distrito y constituye un mecanismo fundamental de gestión ética y responsable.

Sección 1: Información General

Entidad

Nombre de la entidad distrital responsable del sistema de IA.

Unidad responsable

Dependencia interna que gestiona el proyecto.

Título del sistema

Nombre del sistema de IA que se evaluará.

Responsable ARA/DPIA

Persona encargada de liderar la evaluación de riesgos.

Responsable técnico

Funcionario o profesional que entrega el soporte tecnológico.

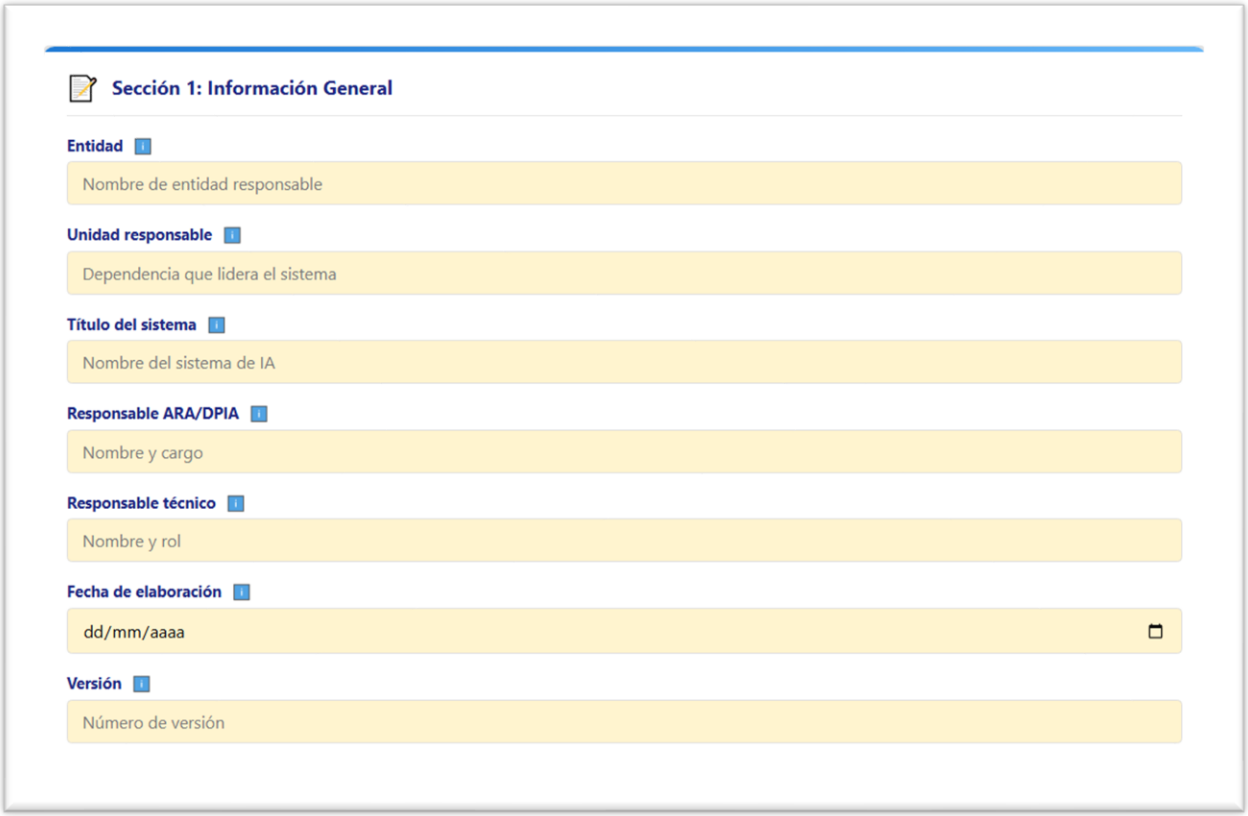
Fecha de elaboración

Fecha en la que se completa el formulario.

Versión

Número de versión del documento (para trazabilidad).

Forma 23. Información General – Formulario ARA/DPIA



Sección 2: Descripción del Sistema y Alcance

Descripción técnica del sistema

Explicación sencilla de cómo funciona y qué hace la IA.

Finalidad específica

Propósito concreto del sistema y problema que resuelve.

Base legal

Norma o documento que justifica su implementación.

Población objetivo

Personas o grupos afectados o beneficiados por la IA.

Casos de uso permitidos

Aplicaciones autorizadas del sistema.

Casos de uso no permitidos

Escenarios en los que su uso está prohibido.

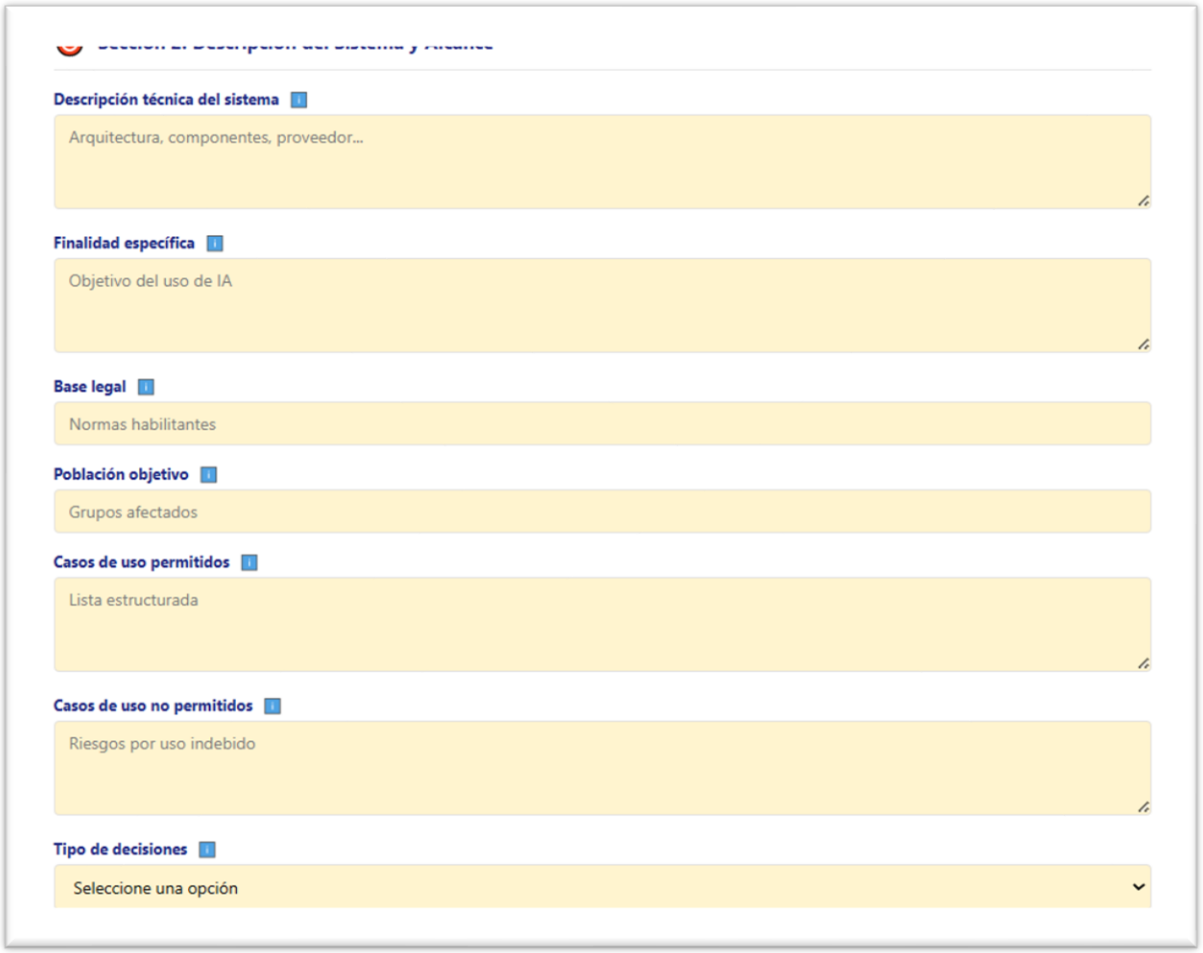
Tipo de decisiones

Indica si la IA informa, recomienda o decide automáticamente.

Contexto de uso

Situaciones o procesos donde se aplicará la IA.

Forma 24. Descripción del sistema y alcance – Formulario ARA/DPIA



Sección 3: Datos y Origen de la Información

3.1 Tabla de Mapeo de Datos

Categoría del dato

Tipo general del dato (identificación, salud, contacto, etc.).

Tipo de dato

Estructura del dato (texto, número, imagen, etc.).

Fuente

De dónde proviene el dato (formularios, bases internas, terceros).

Dato personal (Sí/No)

Si la información identifica o puede identificar a una persona.

Base de licitud

Fundamento legal para usar el dato.

Finalidad

Para qué se usa cada dato dentro del sistema.

Técnicas aplicadas

Métodos usados (análisis, anonimización, limpieza, etc.).

Sesgos identificados

Problemas detectados en representatividad o equilibrio.

Acciones propuestas

Medidas para corregir sesgos o mejorar la calidad del dato.

3.2 Documentación del Dataset

Nombre del dataset

Cómo se identifica el conjunto de datos.

Origen

Quién lo generó o recopiló.

Motivación

Por qué se recolectaron los datos.

Método de recolección

Cómo se obtuvieron los datos.

Procesos de limpieza o anonimización

Ajustes realizados antes de entrenar la IA.

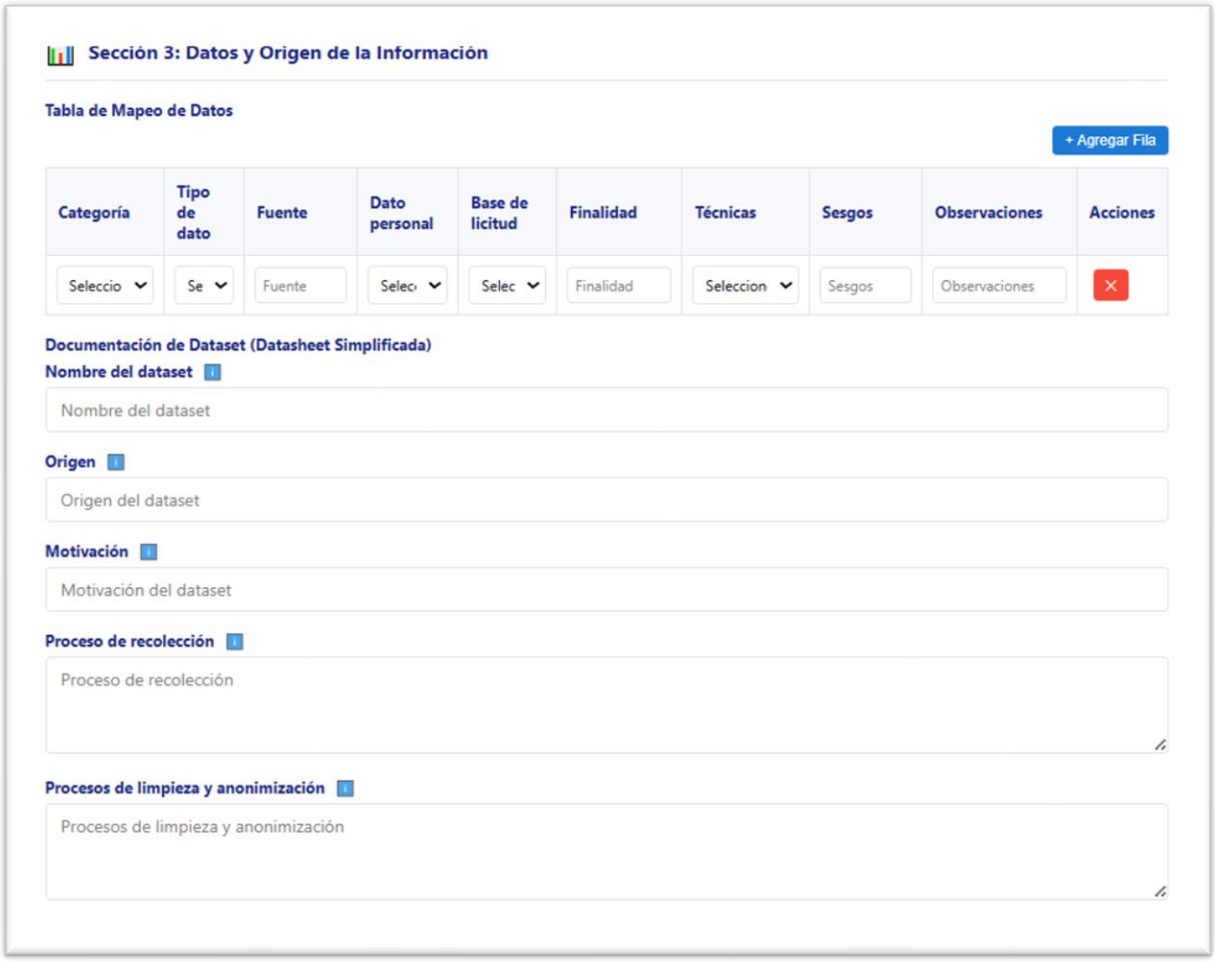
Usos previstos

Aplicaciones permitidas.

Usos no previstos

Aplicaciones prohibidas.

Forma 25. Datos y origen e la información – Formulario ARA/DPIA



Sección 4: Base Legal y Consentimiento

Base legal del tratamiento

Regla que autoriza recolectar y usar datos.

¿Requiere consentimiento?

Indica si los titulares deben autorizar el tratamiento.

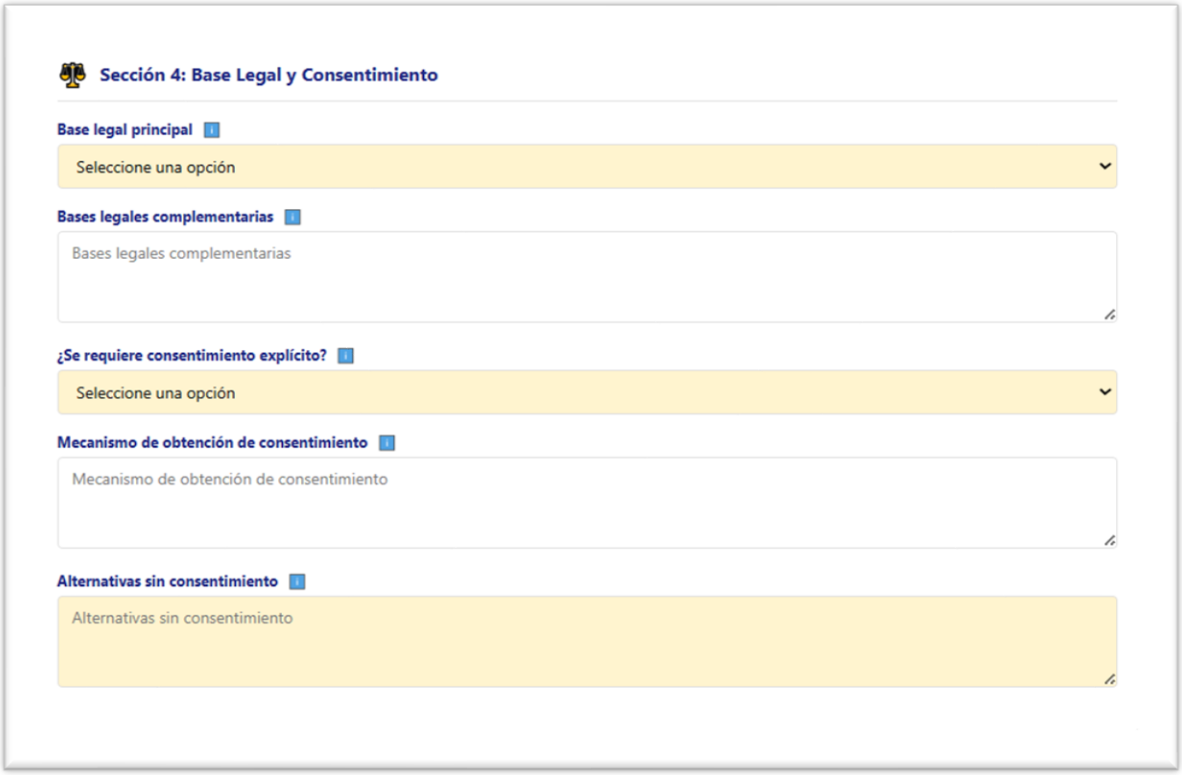
Mecanismo de consentimiento

Cómo se solicita y obtiene la autorización.

Alternativas sin consentimiento

Opciones cuando no se requiere autorización expresa.

Forma 26. Base legal y consentimiento – Formulario ARA/DPIA



Sección 5: Evaluación de Impactos en Derechos

Preguntas por derecho fundamental

Guía para evaluar afectación a derechos.

Respuesta

Explicación breve del riesgo.

Probabilidad (1–3)

Qué tan probable es que ocurra.

Impacto (1–3)

Qué tan grave sería si ocurre.

Riesgo calculado

Resultado automático de probabilidad × impacto.

Acciones propuestas

Medidas para reducir o controlar el riesgo.

Forma 27. Evaluación de impactos en derechos – Formulario ARA/DPIA

Sección 6: Matriz de Riesgos Algorítmicos

Descripción del riesgo

Explicación clara del riesgo técnico u operativo.

Categoría

Tipo de riesgo (equidad, privacidad, seguridad, etc.).

Probabilidad

Baja, media o alta.

Impacto

Bajo, medio o alto.

Nivel de riesgo

Clasificación automática.

Controles existentes

Medidas que ya están implementadas.

Acciones propuestas

Medidas adicionales necesarias.

Forma 28. Matriz de riesgos algorítmicos – Formulario ARA/DPIA

Sección 7: Medidas de Mitigación y Controles

Medida o acción

Qué se hará para reducir el riesgo.

Tipo

Si es una medida técnica (IA) u organizativa (proceso).

Responsable

Quién debe implementarla.

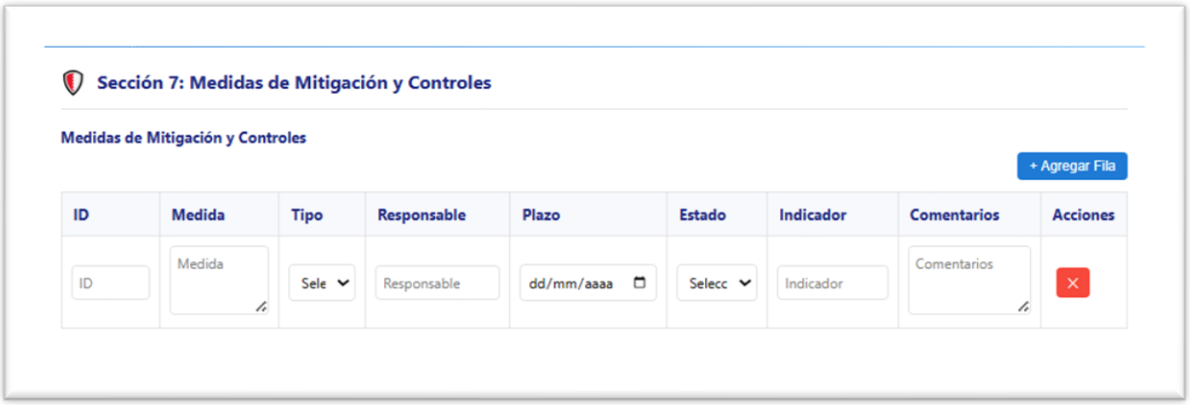
Plazo

Fecha o tiempo estimado para cumplirla.

Estado

Planificada, en curso o implementada.

Forma 29. Medidas de mitigación y controles – Formulario ARA/DPIA



Sección 8: Supervisión Humana

Mecanismos de intervención

Cómo las personas pueden intervenir o anular decisiones de IA.

Criterios de escalamiento

Cuándo un caso debe revisarse manualmente.

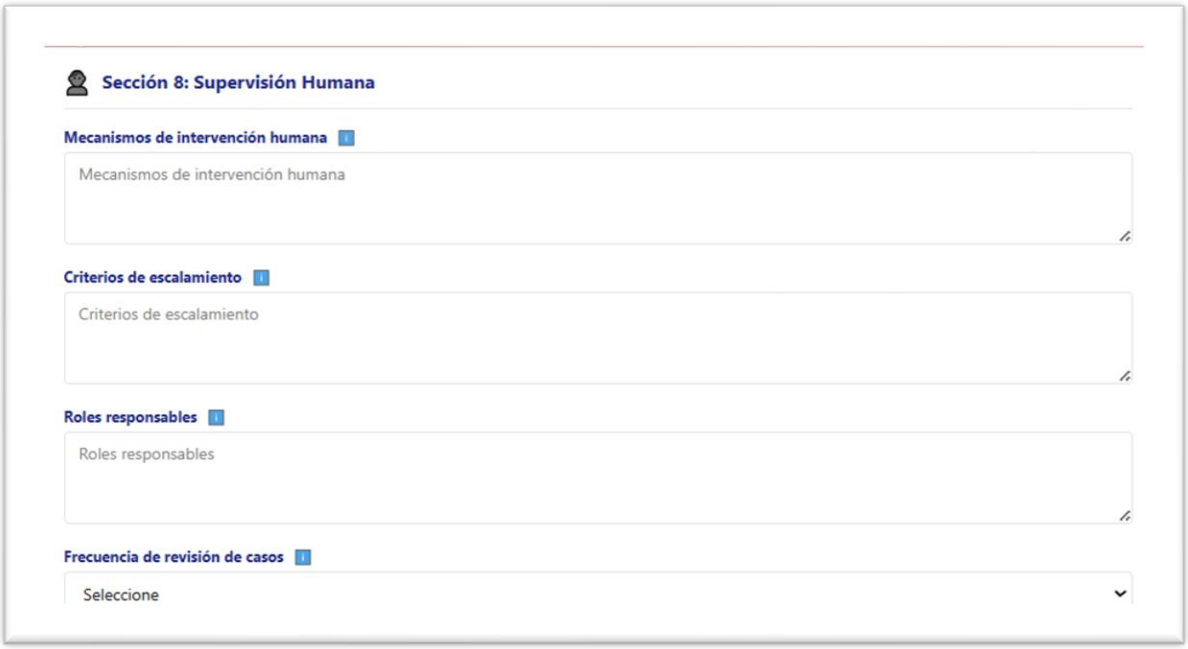
Responsables

Quién supervisa y toma decisiones.

Frecuencia de revisión

Cada cuánto se revisan resultados o errores.

Forma 30. Supervisión humana – Formulario ARA/DPIA



Sección 9: Monitoreo Continuo y KPIs

Categoría del indicador

Si es técnico, de equidad, de privacidad, etc.

Indicador/KPI

Métrica para monitorear desempeño o riesgos.

Valor objetivo

Meta esperada.

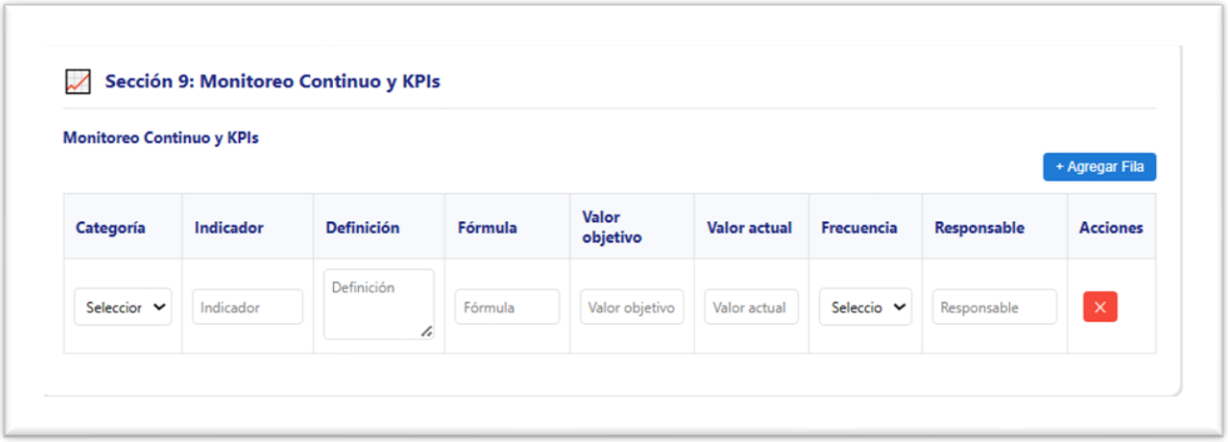
Frecuencia

Mensual, trimestral, anual, etc.

Acciones correctivas

Qué se hace cuando un indicador falla.

Forma 31. Monitoreo continuo y KPIs – Formulario ARA/DPIA



Sección 10: Comunicación y Transparencia

Mensajes a usuarios

Información que se dará a los ciudadanos.

Documentos publicados

Políticas, fichas técnicas, advertencias.

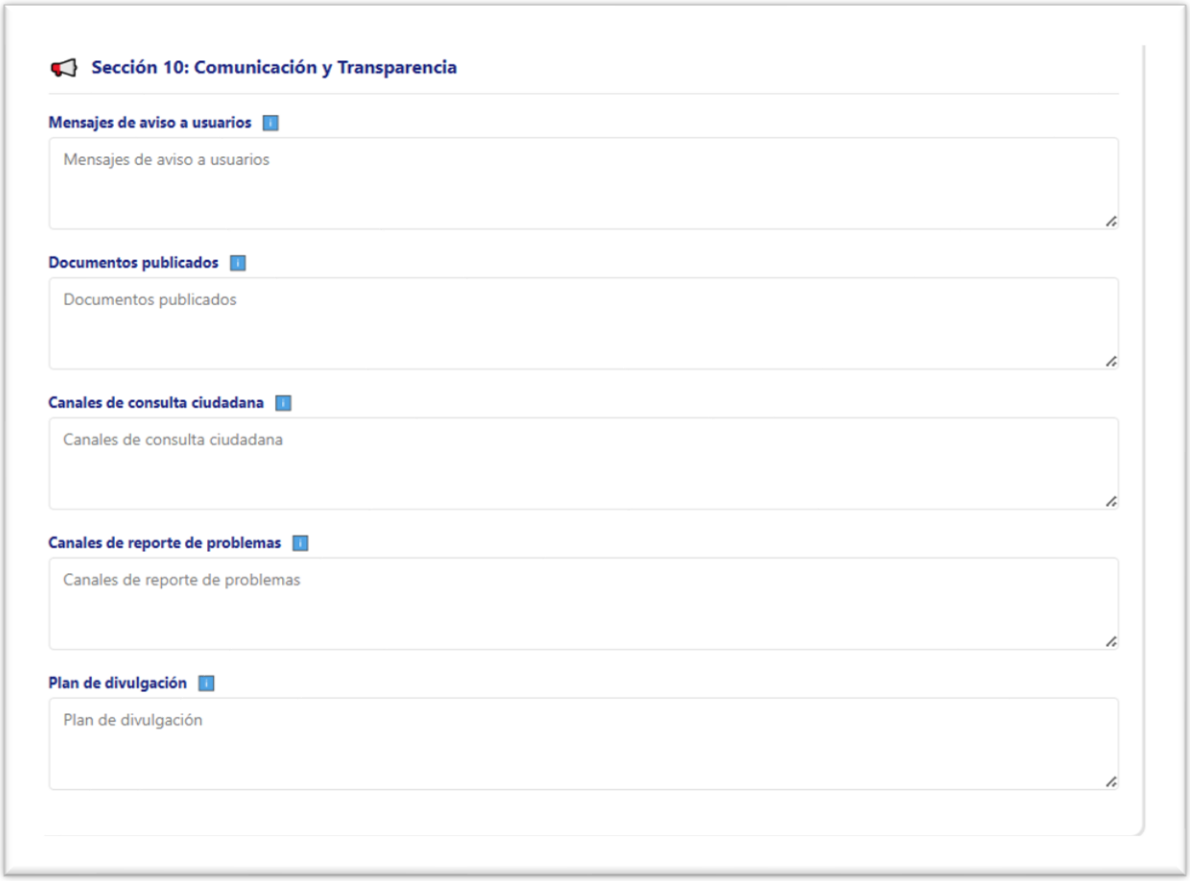
Canales de consulta

Líneas o medios para solicitar información.

Canales de reporte

Cómo denunciar errores o riesgos.

Forma 32. Comunicación y transparencia – Formulario ARA/DPIA



Sección 11: Auditoría y Actualizaciones

Tipo de auditoría

Interna o externa.

Hallazgos

Principales problemas detectados.

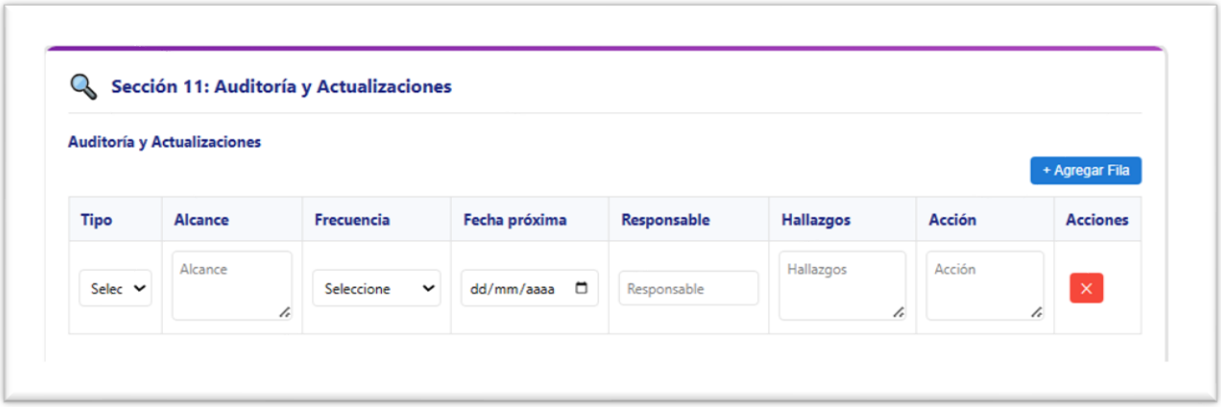
Acciones correctivas

Medidas a implementar.

Frecuencia

Cada cuánto se audita.

Forma 33. Auditoría y actualizaciones – Formulario ARA/DPIA



Sección 12. Aprobaciones y Decisión

Evaluación técnica, jurídica y del DPO

Conceptos de cada dependencia.

Decisión final del Comité de IA

Aprobado, aprobado con condiciones o rechazado.

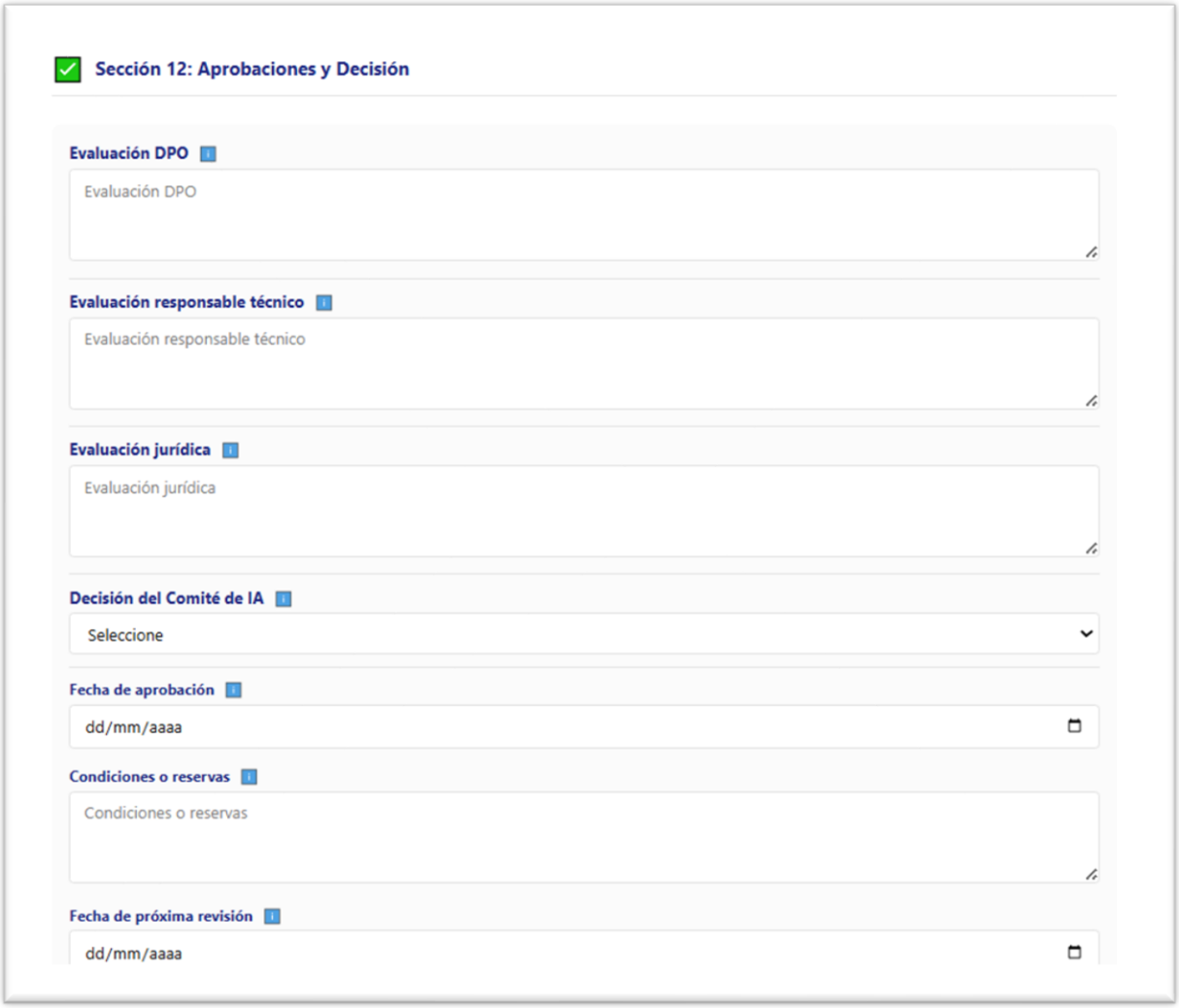
Condiciones

Requisitos que se deben cumplir para avanzar.

Fecha de próxima revisión

Momento en que se debe reevaluar el sistema.

Forma 34. Aprobaciones y decisión – Formulario ARA/DPIA



Anexo F4. Manual Funcional del Toolkit Data Sheets

Documentación Estandarizada de Conjuntos de Datos para Sistemas de IA.

Introducción

El Toolkit Data Sheets es una herramienta del Framework Distrital de Gobernanza de Inteligencia Artificial diseñada para documentar de manera clara, estructurada y transparente los conjuntos de datos utilizados en el desarrollo, entrenamiento, validación y operación de sistemas de IA en entidades públicas del Distrito Capital.

Su objetivo principal es garantizar la calidad, trazabilidad, equidad y uso responsable de los datos, permitiendo identificar riesgos asociados a sesgos, representatividad insuficiente, problemas de privacidad o usos indebidos de la información. Esta herramienta es especialmente relevante en el sector público, donde los datos suelen estar vinculados directamente con derechos fundamentales de la ciudadanía.

El uso del Data Sheet es obligatorio para todo proyecto de IA que utilice datos, y constituye un insumo clave para:

- la Evaluación de Riesgos Algorítmicos (ARA),
- la Evaluación de Impacto en Protección de Datos (DPIA),
- la auditoría y supervisión de sistemas de IA,

la rendición de cuentas ante entes de control y ciudadanía.

Sección 1: Información General del Dataset

Estos campos sirven para identificar el dataset y saber quién lo administra.

- **Nombre del dataset**

  Nombre claro para identificar este conjunto de datos.

- **Entidad propietaria**

  La entidad del Distrito que posee los datos.

- **Área responsable**

  La dependencia o unidad interna que cuida y administra los datos.

- **Versión**

  Número o código que permite saber si el dataset ha cambiado con el tiempo.

- **Fecha de creación**

  El día en que se creó este conjunto de datos.

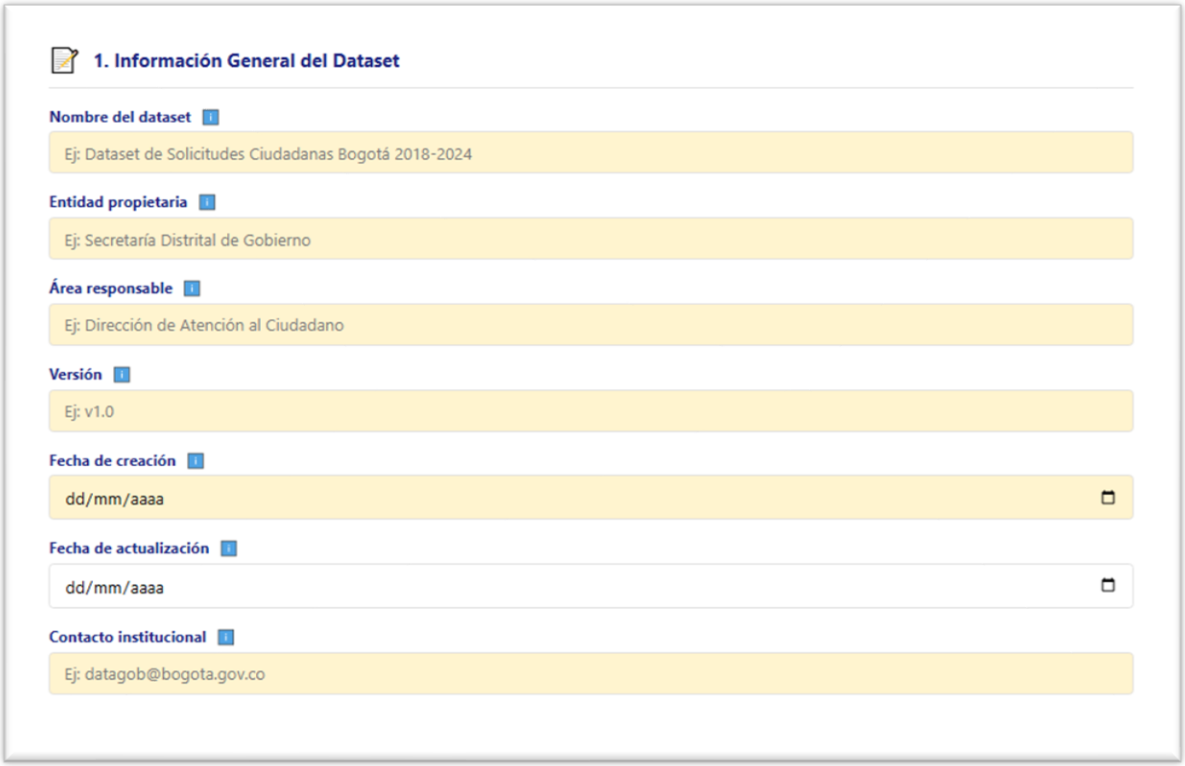
- **Fecha de actualización**

  Cuándo fue la última vez que se actualizó.

- **Contacto institucional**

  Información para comunicarse con la persona o área responsable.

Forma 35. Información General del dataset – Formulario Data Sheets



Sección 2: Descripción y Propósito

Esta sección describe de qué tratan los datos y para qué se usan.

- **Descripción del contenido**

  Explica qué tipo de información contiene el dataset.

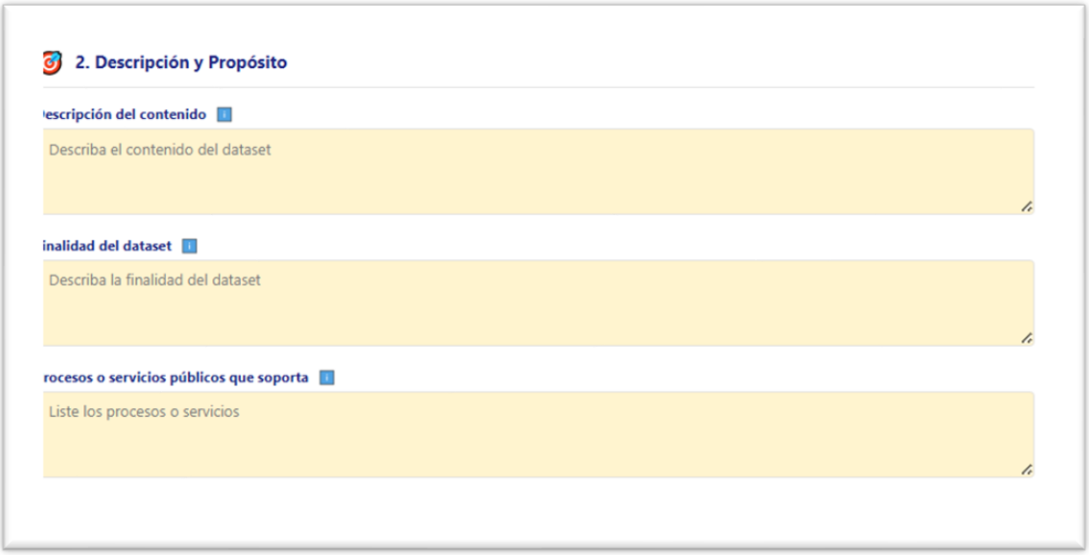
- **Finalidad del dataset**

  Para qué se creó el dataset o qué problema ayuda a resolver.

- **Procesos o servicios públicos que soporta**

  Qué servicios del Distrito utilizan esta información.

Forma 36. Descripción y propósito – Formulario Data Sheets



Sección 3: Origen y Método de Recolección

Aquí se explica de dónde vienen los datos y cómo se obtuvieron.

- **Fuente de los datos**

  Indica de qué sistema o registro provienen.

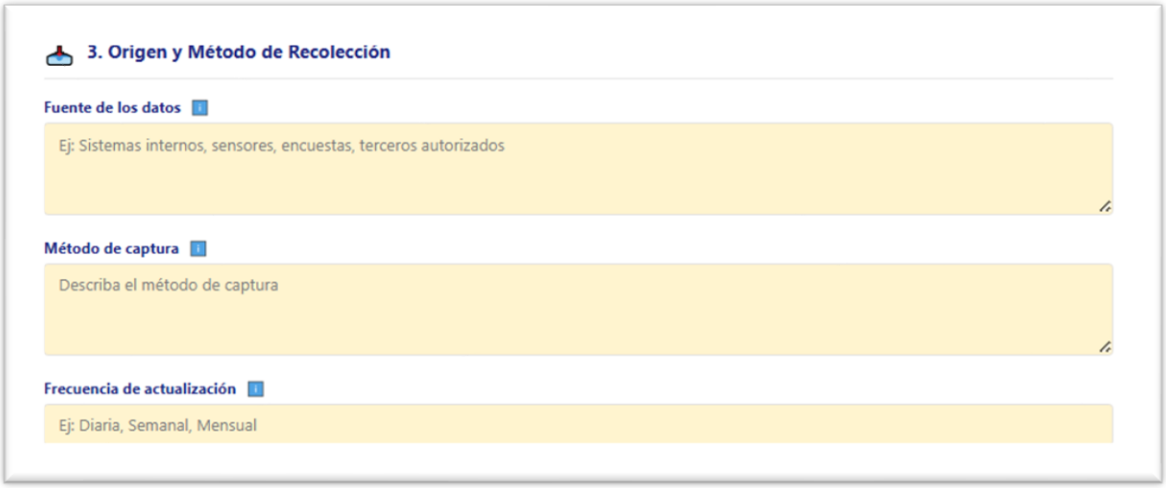
- **Método de captura**

  Cómo se recolectaron (manual, automático, sensores, formularios).

- **Frecuencia de actualización**

  Cada cuánto se actualizan los datos.

Forma 37. Origen y método de recolección – Formulario Data Sheets



Sección 4: Composición del Dataset

Describe la estructura y los contenidos del dataset en detalle.

- **Número de registros**

  Cuántas filas o elementos contiene el dataset.

- **Número de variables**

  Cuántas columnas o campos se registran.

- **Tipos de datos**

  Qué tipo de información hay (texto, número, fecha, etc.).

- **Representatividad poblacional**

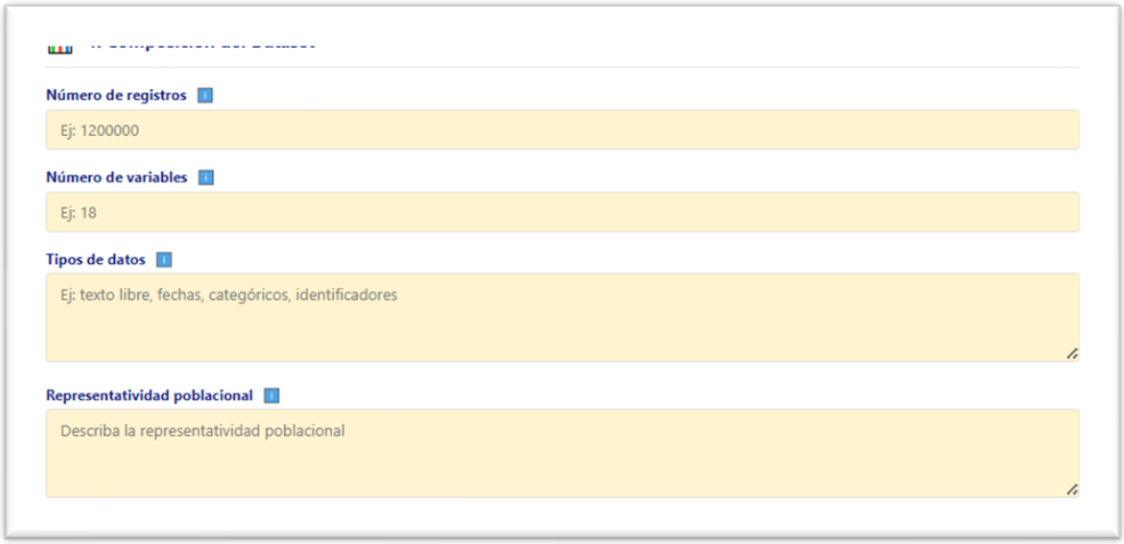
- Si los datos representan a toda la población o solo a parte.

Además, incluye un **diccionario de datos** con estos campos por variable:

- Nombre de la variable
- Descripción
- Tipo de dato
- Valores permitidos o rango
- ¿Dato personal?

- ¿Dato sensible?
- ¿Es obligatorio?
- Observaciones
- Acciones sugeridas

Forma 38. Composición del dataset – Formulario Data Sheets



Sección 5: Presencia de Datos Personales y Sensibles

Evalúa si el dataset contiene información que identifica personas o es delicada.

- **Identificación del tipo de datos personales/sensibles**

  Enumera qué datos pueden identificar a una persona o son delicados.

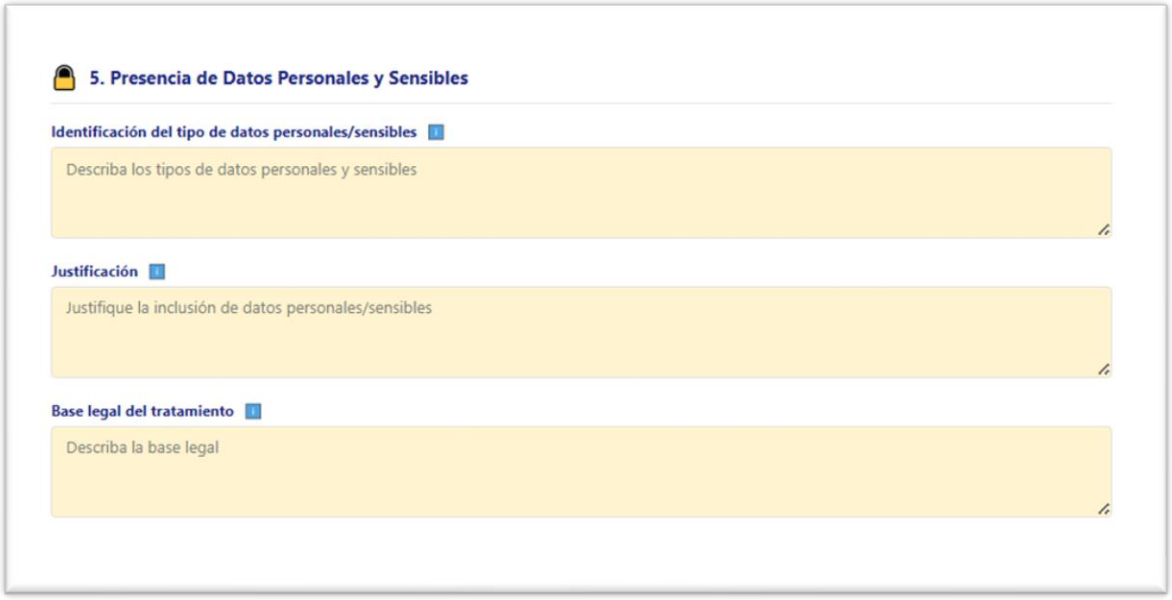
- **Justificación**

  Por qué se incluyen esos datos.

- **Base legal del tratamiento**

  Norma que permite usar estos datos (por ejemplo, Ley 1581 de 2012).

Forma 39. Presencia de datos personales y sensibles – Formulario Data Sheets



Sección 6: Calidad del Dataset

Evalúa qué tan útiles y completos son los datos.

- **Datos faltantes**

  Cuántos datos hacen falta.

- **Inconsistencias**

  Problemas de error entre datos.

- **Procesos de limpieza aplicados**

  Qué se hizo para corregir problemas.

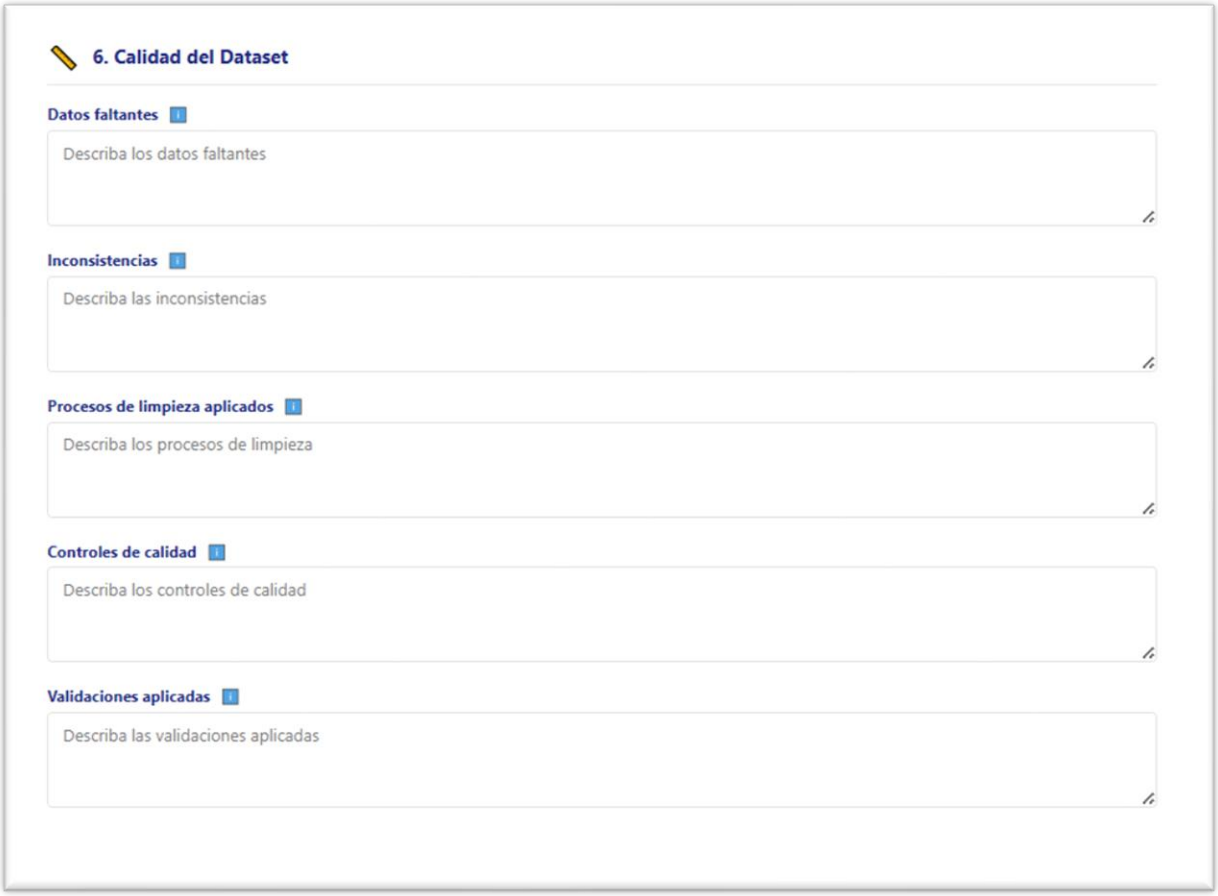
- **Controles de calidad**

  Procedimientos para asegurar que los datos sean confiables.

- **Validaciones aplicadas**

  Revisiones que confirman que los datos sean correctos.

Forma 40. Calidad del dataset – Formulario Data Sheets



Sección 7: Evaluación de Sesgos y Representatividad

Analiza si los datos son equitativos y representan adecuadamente a la población.

- **Sesgos identificados**

  Descripción de posibles desviaciones injustas (por género, edad, etc.).

- **Distribución de variables sensibles**

  Cómo se distribuyen grupos sensibles en los datos.

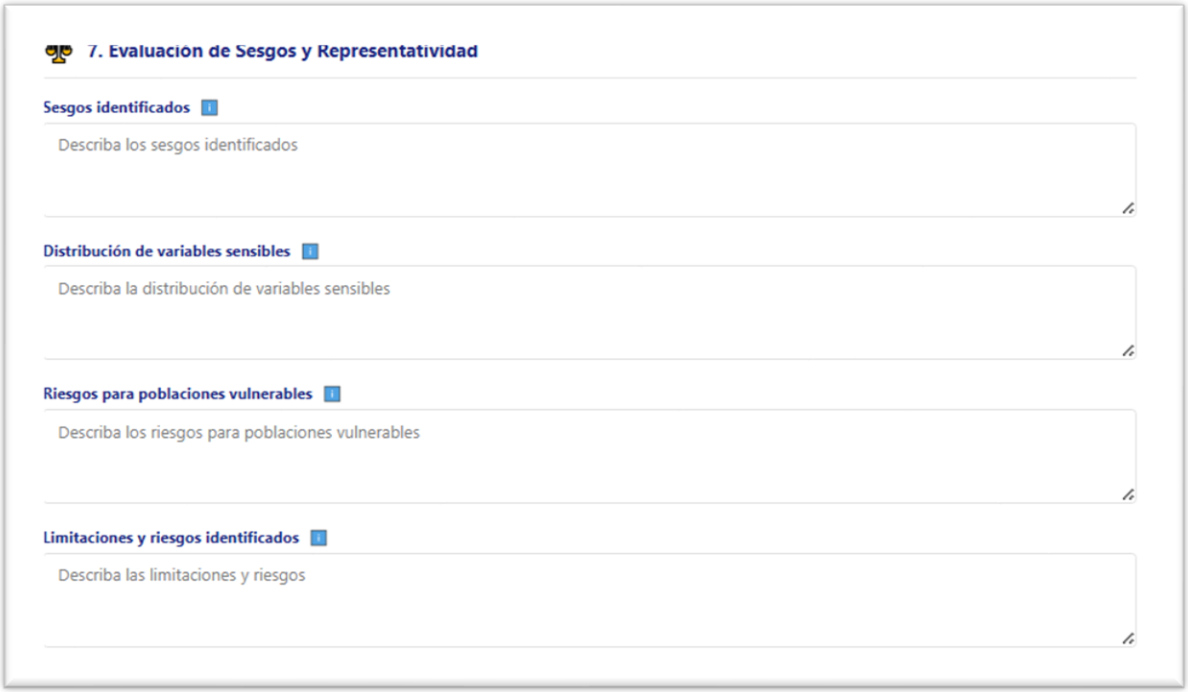
- **Riesgos para poblaciones vulnerables**

  Qué grupos podrían verse perjudicados por sesgos.

- **Limitaciones y riesgos identificados**

  Otras restricciones que afectan la representatividad.

Forma 41. Evaluación de sesgos y representatividad – Formulario Data Sheets



Sección 8: Procesamiento y Transformaciones Aplicadas

Describe qué se hizo para preparar los datos antes de usar.

- **Normalización**

  Ajustes para que las variables estén en un formato estándar.

- **Imputación**

  Cómo se completaron datos que estaban faltantes.

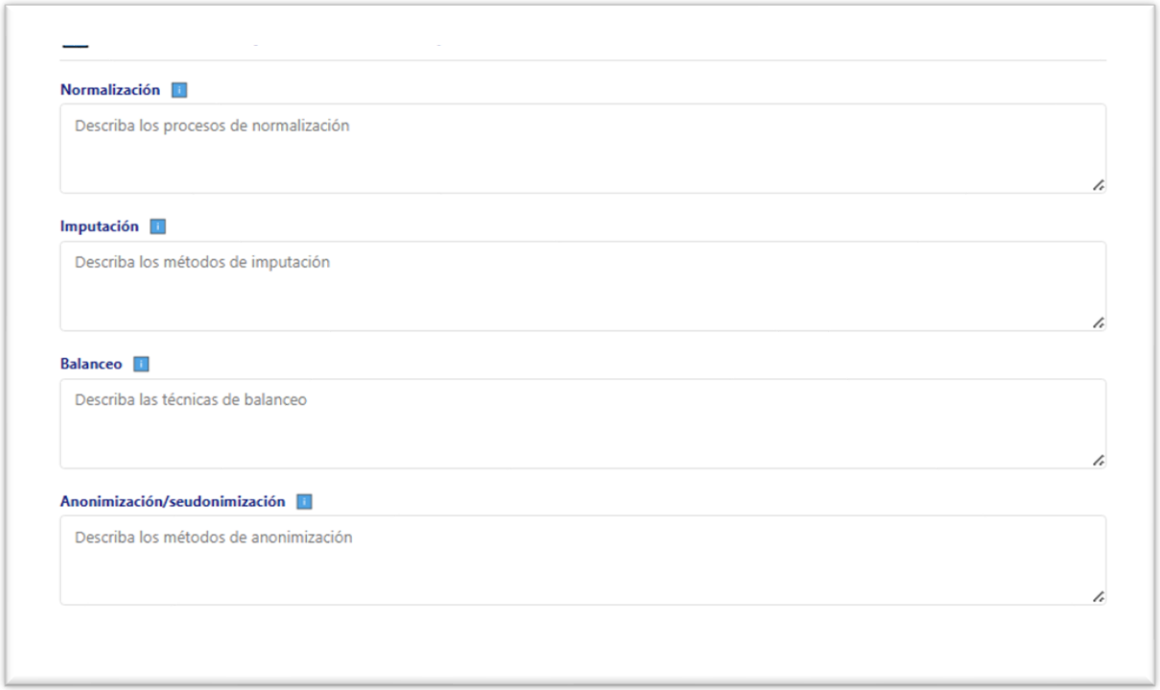
- **Balanceo**

  Ajustes para equilibrar la representación de grupos.

- **Anonimización/seudonimización**

  Procesos para proteger la identidad de las personas.

Forma 42. Procesamiento y transformaciones aplicadas – Formulario Data Sheets



Sección 9: Riesgos Éticos y Legales

Analiza riesgos relacionados con discriminación o uso indebido.

- **Riesgos de discriminación**

  Si los datos pueden llevar a resultados injustos.

- **Riesgos de reidentificación**

  Posibilidad de identificar personas aunque los datos estén anonimizado.

- **Riesgos de uso indebido**

  Posibles usos malintencionados.

- **Riesgos de sesgos estructurales**

  Problemas profundos de representación.

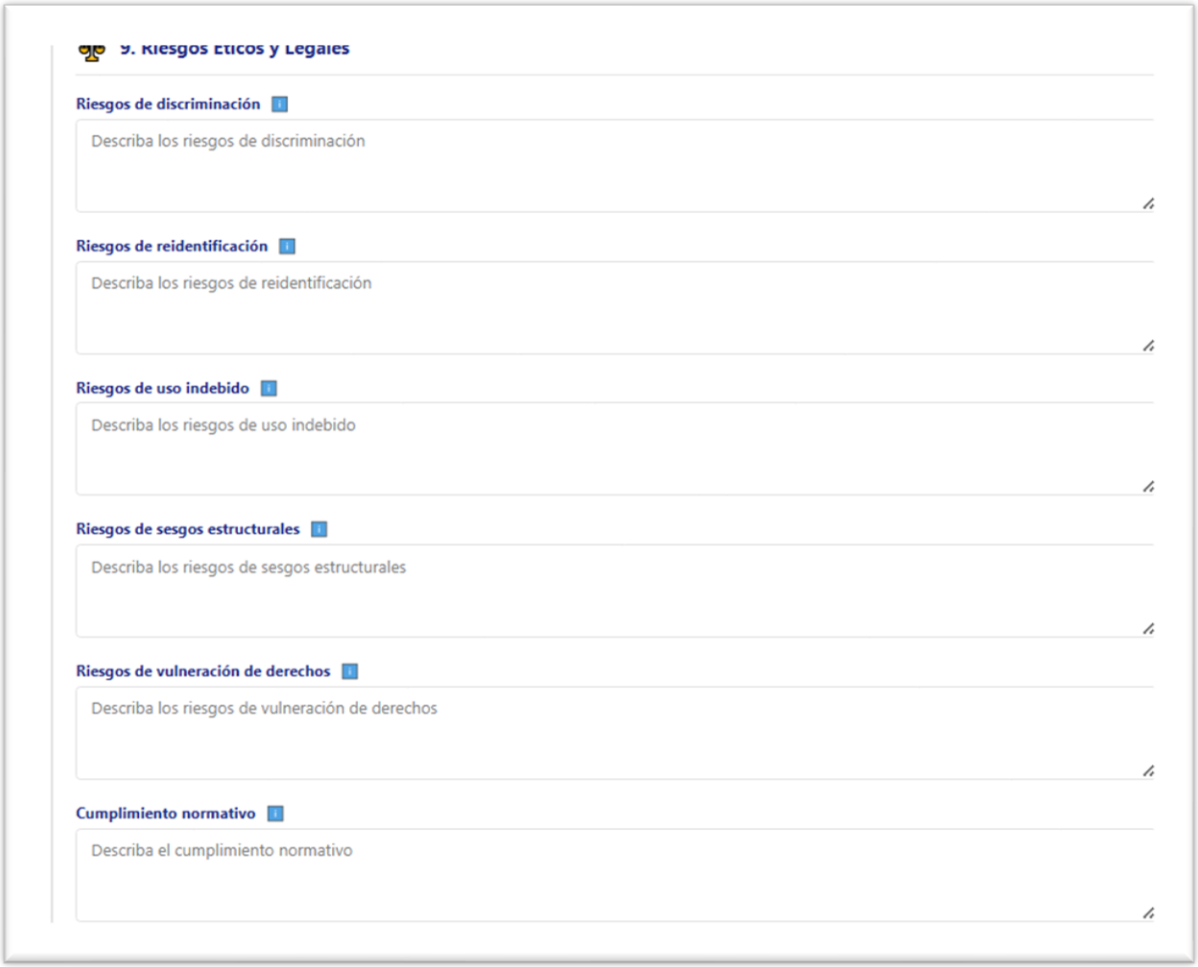
- **Riesgos de vulneración de derechos**

  Posibles afectaciones a derechos fundamentales.

- **Cumplimiento normativo**

  Si se ajusta a las leyes vigentes.

Forma 43. Riesgos éticos y legales



Sección 10: Uso Permitido y No Permitido

Define claramente cómo se pueden usar (o no) los datos.

- **Finalidad autorizada**

  Para qué sí se pueden usar los datos.

- **Restricciones**

  Lo que está prohibido.

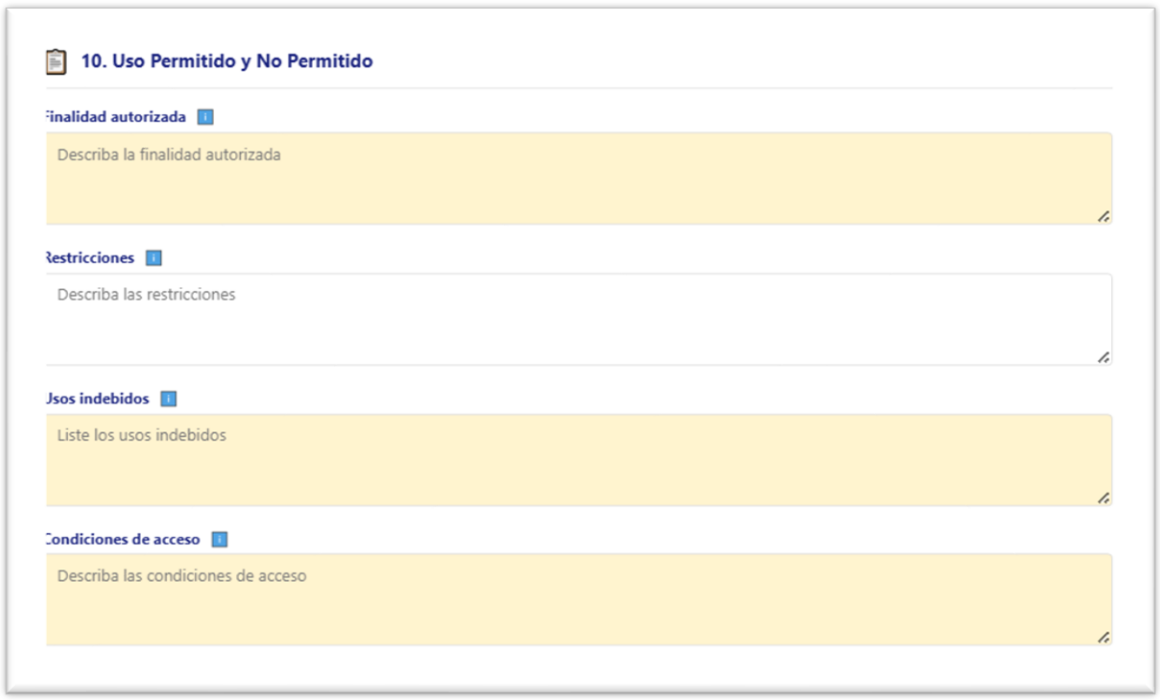
- **Usos indebidos**

  Ejemplos de usos que no deben hacerse.

- **Condiciones de acceso**

  Quién, cómo y bajo qué condiciones puede acceder.

Forma 44. Uso permitido y no permitido – Formulario Data Sheets



Sección 11: Seguridad del Dataset

Describe las medidas que protegen los datos.

- **Controles de seguridad**

  Herramientas y procedimientos para proteger los datos.

- **Cifrado**

  Si los datos están codificados para protegerlos.

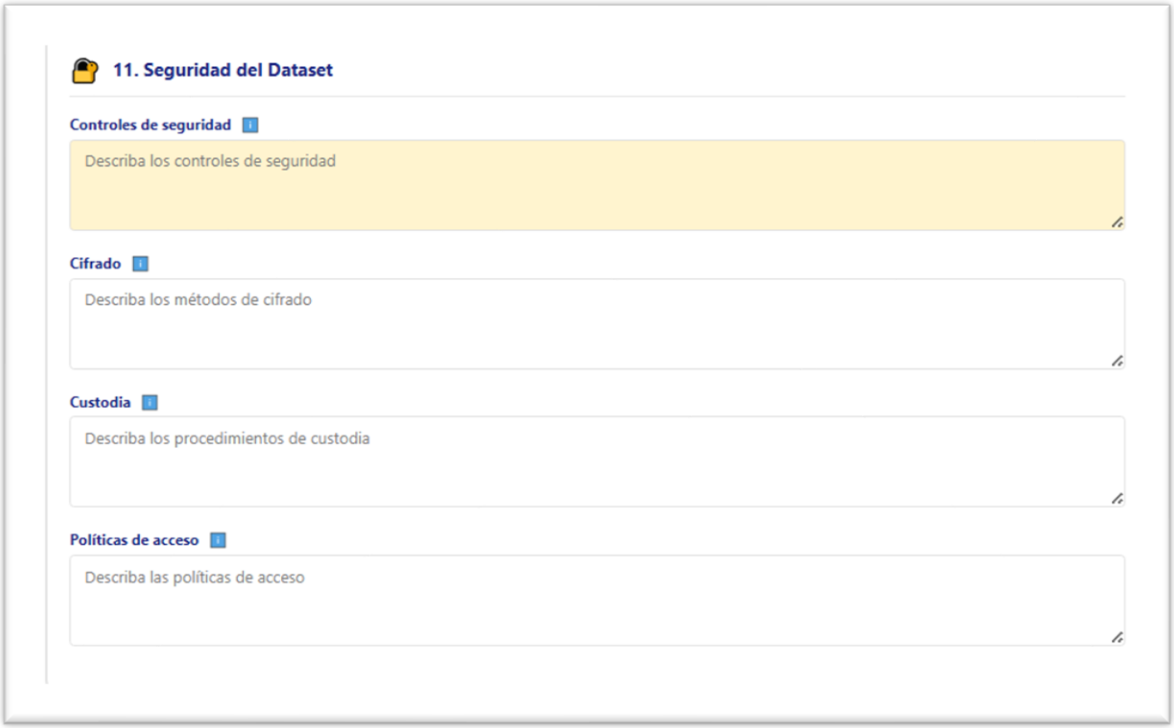
- **Custodia**

  Quién guarda y administra los datos.

- **Políticas de acceso**

  Reglas sobre quién puede ver o usar el dataset.

Forma 45. Seguridad del dataset – Formulario Data Sheets



Sección 12: Historial del Dataset

Registra los cambios hechos con el tiempo.

Cada fila registra:

- **Versión**

  Número de actualización.

- **Fecha**

  Cuándo se hizo el cambio.

- **Cambios realizados**

  Qué se modificó.

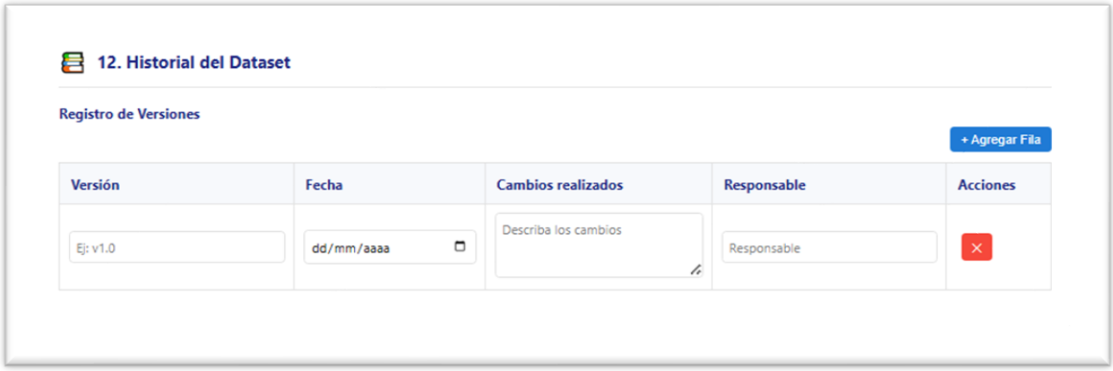
- **Responsable**

  Quién hizo el cambio.

- **Acciones**

  Qué se hizo para aprobar o documentar el cambio.

Forma 46. Historial del Dataset - Formulario Data Sheets



Sección 13: Aprobaciones Institucionales

Indica quiénes revisaron y aprobaron el dataset.

- **Equipo de datos**

  Validación técnica.

- **Área jurídica**

  Verificación legal.

- **Comité Distrital de IA**

  Aprobación final institucional.

Forma 47. Aprobaciones institucionales – Formulario Data Sheets

Anexo F5. Manual Funcional para el Checklist de Evaluación de Proveedores de IA

Instrumento estandarizado para la debida diligencia y selección responsable de proveedores de soluciones, servicios o productos basados en IA.

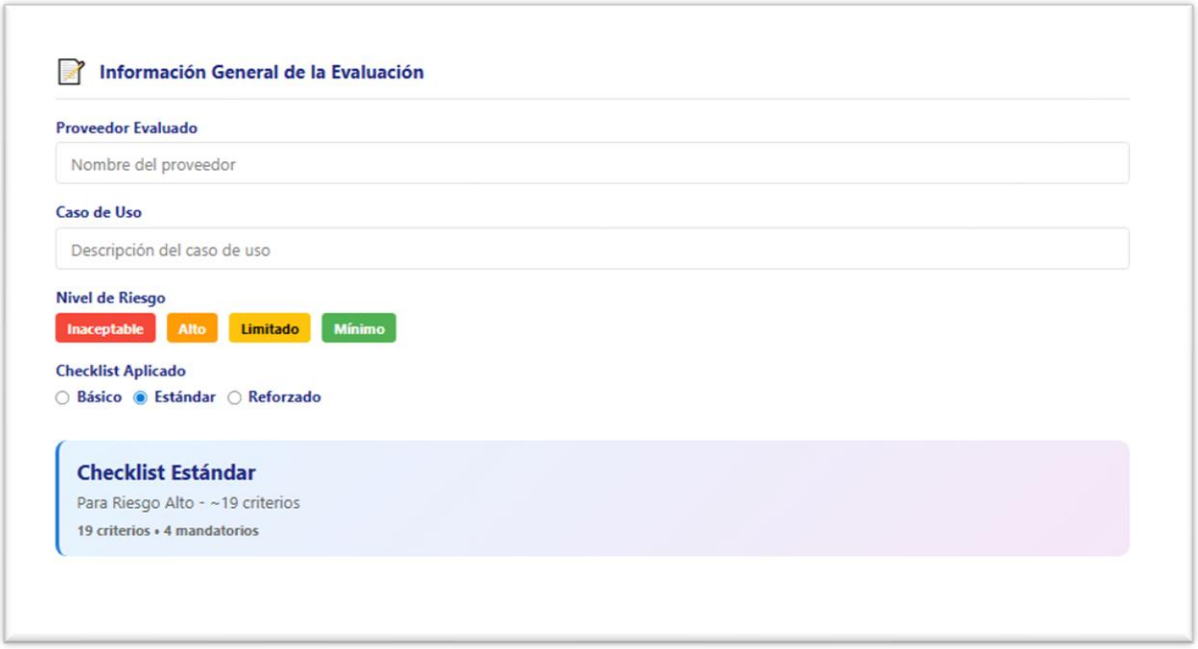
Introducción a la Herramienta

El Checklist de Evaluación de Proveedores de IA es una herramienta del Framework de Gobernanza de Inteligencia Artificial del Distrito Capital que permite a las entidades públicas evaluar de manera estructurada, objetiva y transparente a los proveedores de soluciones basadas en inteligencia artificial antes de su contratación o adopción.

Su finalidad es reducir riesgos legales, éticos, técnicos y operativos, fortalecer la posición contractual de la entidad y asegurar que los proveedores cumplan con estándares mínimos de cumplimiento normativo colombiano, protección de datos personales, seguridad de la información, transparencia técnica y calidad del servicio.

Este checklist no pretende excluir proveedores por falta de madurez internacional, sino establecer una línea base razonable y progresiva, acorde con el contexto colombiano y distrital, que permita iniciar procesos de adopción responsable de IA en el sector público.

Forma 48. Información general de Evaluación IA – Checklist Proveedores



1. Conformidad Regulatoria y Gobernanza

Esta sección del checklist evalúa si el proveedor de soluciones de inteligencia artificial cuenta con las condiciones mínimas de cumplimiento normativo, principios éticos y estructuras básicas de gobernanza, necesarias para operar de manera responsable con entidades públicas del Distrito Capital de Bogotá.

El objetivo no es exigir niveles avanzados de madurez internacional, sino verificar que el proveedor esté preparado para cumplir la normativa colombiana vigente y alinearse progresivamente con las políticas públicas de IA.

1.1 Conformidad con el marco regulatorio colombiano

En este criterio se evalúa si el proveedor demuestra conocimiento, comprensión y cumplimiento del marco legal colombiano aplicable al tratamiento de datos y al uso de tecnologías digitales en el sector público.

El evaluador debe verificar si el proveedor:

- Conoce y cumple la Ley 1581 de 2012 sobre protección de datos personales.
- Aplica las disposiciones y lineamientos emitidos por la Superintendencia de Industria y Comercio (SIC).
- Cuenta con una política de tratamiento de datos personales pública, vigente y accesible.

- Puede comprometerse contractualmente a cumplir los lineamientos de política pública de IA definidos en el CONPES 4144 y demás directrices aplicables al Distrito Capital.

Este criterio es obligatorio: una calificación de "0" implica la no viabilidad del proveedor para procesos contractuales con la entidad.

1.2 Política de IA Responsable o Ética

Este criterio evalúa si el proveedor ha definido principios, lineamientos o compromisos explícitos para el uso ético y responsable de la inteligencia artificial.

El evaluador debe revisar si el proveedor:

- Cuenta con una política de IA responsable, ética o uso responsable de tecnologías de IA, ya sea pública o interna.
- Declara principios relacionados con equidad, no discriminación, transparencia, supervisión humana o responsabilidad.
- Ha establecido mecanismos internos (formales o informales) para revisar el impacto ético de sus soluciones.

No se exige un modelo avanzado, sino una base mínima que permita a la entidad distrital exigir coherencia ética durante la ejecución del contrato.

1.3 Sistema de Gestión de IA (AIMS)

Este criterio analiza si el proveedor dispone de un sistema estructurado —formal o en desarrollo— para gestionar soluciones de IA a lo largo de su ciclo de vida.

El evaluador debe verificar si el proveedor:

- Cuenta con un Sistema de Gestión de IA (AIMS) documentado, o
- Dispone de políticas, procedimientos o controles internos que regulen el diseño, despliegue, operación y mejora de sistemas de IA.

La certificación en normas internacionales (por ejemplo, ISO/IEC 42001) no es obligatoria, pero puede valorarse positivamente si existe evidencia equivalente de gestión sistemática.

1.4 Gestión de Riesgos de IA

Este criterio evalúa si el proveedor reconoce que los sistemas de IA generan riesgos y si cuenta con mecanismos básicos para identificarlos, evaluarlos y mitigarlos.

El evaluador debe revisar si el proveedor:

- Identifica riesgos asociados a privacidad, sesgos, seguridad, continuidad del servicio o reputación.
- Aplica metodologías básicas de análisis de riesgos en sus soluciones de IA.
- Puede documentar acciones de mitigación o controles implementados frente a riesgos identificados.

Este criterio es clave para asegurar que el proveedor pueda articularse con la Matriz de Riesgos de IA y el ARA/DPIA del framework distrital.

Resultado de la Sección 1

La calificación de esta sección permite a la entidad determinar si el proveedor cuenta con las condiciones mínimas de legalidad, ética y gobernanza para iniciar una relación contractual en proyectos de inteligencia artificial.

Una baja calificación no implica necesariamente exclusión inmediata, pero sí la necesidad de establecer compromisos contractuales, planes de mejora o condiciones previas a la adjudicación.

Forma 49. Conformidad regulatoria y gobernanza – Checklist Proveedores

SECCIÓN 1: CONFORMIDAD REGULATORIA Y GOBERNANZA (Peso: 25%)

ID	Criterio de Evaluación	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia/Referencia
1.1	**Conformidad con el marco regulatorio colombiano (Ley 1581 de 2012, SIC).**	No tiene política de privacidad o desconoce la normativa.	Tiene política de privacidad básica pero no totalmente alineada con la Ley 1581. No demuestra conocimiento profundo.	Política de privacidad robusta, actualizada y claramente alineada con la Ley 1581. Compromiso contractual explícito.	0	Documento o referencia
1.2	**Política de IA Responsable o Ética.**	No tiene política documentada.	Tiene un borrador o política interna, pero no es pública ni está formalizada.	Política pública formal y referenciable (ej.: en su sitio web), con principios claros y mecanismos de implementación.	0	Documento o referencia
1.3	**Sistema de Gestión de IA (AIMS).**	No tiene un sistema de gestión para IA.	Tiene elementos de un sistema (políticas, roles) pero no está formalizado, documentado o es inconsistente.	Tiene un sistema documentado e implementado (ej.: basado en ISO/IEC 42001) o certificación ISO/IEC 42001 vigente.	0	Documento o referencia
1.4	**Gestión de Riesgos de IA.**	No tiene un proceso definido para identificar y gestionar riesgos de IA.	Realiza evaluaciones de riesgo de manera ad-hoc o solo para proyectos específicos, sin un proceso sistemático.	Tiene un proceso sistemático y documentado de gestión de riesgos de IA.	0	Documento o referencia

CALIFICACIÓN SECCIÓN 1: 0/8

2. Documentación y Transparencia Técnica

Esta sección del checklist evalúa el nivel de transparencia técnica y documental que ofrece el proveedor respecto a la solución de inteligencia artificial propuesta. Su objetivo es asegurar que la entidad pública pueda comprender, supervisar, auditar y explicar el funcionamiento del sistema, tanto a nivel interno como frente a organismos de control y a la ciudadanía.

En el contexto del Distrito Capital de Bogotá, esta sección es clave para garantizar los principios de transparencia, trazabilidad y rendición de cuentas en el uso de tecnologías algorítmicas.

2.1 Model Card (Ficha del Modelo)

Este criterio evalúa si el proveedor entrega una Ficha del Modelo (Model Card) que documente de manera clara y estructurada el sistema de IA ofrecido.

El evaluador debe verificar si el proveedor proporciona información sobre:

- La finalidad del modelo y los casos de uso previstos.
- Los casos de uso no previstos o no recomendados.
- El tipo de modelo utilizado (por ejemplo, clasificación, predicción, recomendación).
- Métricas generales de desempeño del modelo.
- Principales limitaciones técnicas y riesgos conocidos.

La Model Card permite a la entidad comprender qué hace el modelo, para qué sirve y cuáles son sus límites, facilitando decisiones informadas y una supervisión responsable.

2.2 Data Sheet (Ficha de Datos)

Este criterio evalúa si el proveedor entrega una Ficha de Datos (Data Sheet) que documente los conjuntos de datos utilizados para entrenar, validar o operar el sistema de IA.

El evaluador debe revisar si el proveedor documenta:

- El origen de los datos utilizados.
- El método general de recolección de la información.
- La composición básica del conjunto de datos.
- La existencia de datos personales o sensibles, cuando aplique.
- Posibles sesgos, limitaciones o problemas de representatividad conocidos.

La Data Sheet es fundamental para evaluar riesgos asociados a privacidad, equidad y calidad de los datos, especialmente en sistemas que impactan a la ciudadanía.

2.3 Documentación Técnica para Auditoría

Este criterio evalúa si el proveedor puede suministrar documentación técnica suficiente para permitir auditorías técnicas, legales o administrativas, sin necesidad de revelar información protegida innecesariamente.

El evaluador debe verificar si el proveedor puede proporcionar:

- Documentación sobre la arquitectura general del sistema.
- Descripción del proceso de entrenamiento o conFormación del modelo.

- Información sobre versiones, cambios relevantes y actualizaciones.
- Evidencias que permitan reconstruir o explicar el funcionamiento del sistema ante requerimientos de control.

Este criterio no exige acceso irrestricto al código fuente, pero sí un nivel de documentación que permita trazabilidad y rendición de cuentas, acorde con las obligaciones del sector público.

Resultado de la Sección 2

La calificación de esta sección permite determinar si el proveedor ofrece un nivel de documentación y transparencia suficiente para operar el sistema de IA en un entorno público, donde la explicabilidad y la supervisión son requisitos fundamentales.

Una calificación baja indica riesgos elevados en términos de opacidad técnica, dificultad de auditoría y limitaciones para responder ante entes de control, lo cual puede derivar en condiciones contractuales adicionales o en la no recomendación del proveedor.

Forma 50. Documentación y transparencia técnica – Checklist Proveedores

SECCIÓN 2: DOCUMENTACIÓN Y TRANSPARENCIA TÉCNICA (Peso: 20%)

ID	Criterio de Evaluación	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia/Referencia
2.1	**Model Card (Ficha del Modelo).**	No proporciona Model Card o la documentación es insuficiente.	Proporciona un Model Card básico, pero le falta información clave (ej.: métricas desagregadas, limitaciones).	Proporciona un Model Card completo con métricas de desempeño, casos de uso, limitaciones y evaluación de equidad.	0	Documento o referencia
2.2	**Data Sheet (Ficha de Datos).**	No proporciona Data Sheet para los conjuntos de datos de entrenamiento.	Proporciona una descripción básica de los datos, pero sin detalles sobre composición, sesgos o proceso de recolección.	Proporciona un Data Sheet detallado con información sobre recolección, composición, pre-procesamiento, sesgos y limitaciones.	0	Documento o referencia
2.3	**Documentación Técnica para Auditoría.**	No proporciona documentación técnica accesible.	Proporciona documentación técnica superficial o desorganizada.	Proporciona documentación técnica detallada (arquitectura, hiperparámetros, proceso de entrenamiento) y bitácoras de decisiones de diseño.	0	Documento o referencia

CALIFICACIÓN SECCIÓN 2: 0/6

3. Privacidad y Protección de Datos

Esta sección del checklist evalúa si el proveedor de soluciones de inteligencia artificial cuenta con condiciones mínimas y verificables para garantizar la protección de datos personales, el respeto por los derechos de los titulares y el cumplimiento de la normativa colombiana en materia de habeas data.

Dado que muchas soluciones de IA en el sector público utilizan información de la ciudadanía, esta sección es crítica para prevenir riesgos legales, sancionatorios y reputacionales para las entidades del Distrito Capital.

3.1 Claridad sobre la Titularidad de los Datos

Este criterio evalúa si el proveedor define de manera clara y contractual la titularidad y el control sobre los distintos tipos de datos utilizados en la solución de IA.

El evaluador debe verificar si el proveedor:

- Distingue entre datos suministrados por la entidad, datos propios del proveedor y datos generados por el sistema.
- Define quién es el responsable del tratamiento y quién actúa como encargado, conforme a la Ley 1581 de 2012.
- Establece claramente los derechos de la entidad sobre los datos durante y después de la ejecución del contrato.

La falta de claridad en este aspecto puede generar conflictos legales y pérdida de control institucional sobre la información pública.

3.2 Data Processing Agreement (DPA)

Este criterio evalúa si el proveedor ofrece un Acuerdo de Tratamiento de Datos Personales (Data Processing Agreement – DPA) conforme a la normativa colombiana.

El evaluador debe revisar si el DPA:

- Define de manera explícita las obligaciones del proveedor como encargado del tratamiento.
- Establece medidas de seguridad técnicas y organizativas para proteger los datos.
- Regula el uso de subencargados, si aplica.
- Incluye obligaciones de notificación en caso de incidentes o brechas de seguridad.

Este criterio es obligatorio: la ausencia de un DPA adecuado descalifica al proveedor para su contratación.

3.3 Gestión de Derechos de los Titulares (Habeas Data)

Este criterio evalúa si el proveedor facilita el ejercicio efectivo de los derechos de los titulares de los datos personales, conforme a la legislación colombiana.

El evaluador debe verificar si el proveedor:

- Cuenta con procedimientos claros para atender solicitudes de acceso, rectificación, actualización, supresión u oposición.
- Puede apoyar técnicamente a la entidad para responder a solicitudes de habeas data dentro de los plazos legales.
- Permite la eliminación o anonimización de datos cuando así lo exija la normativa o la autoridad competente.

Este criterio es especialmente relevante en sistemas de IA que operan de manera continua o que reutilizan datos para aprendizaje.

3.4 Políticas de Retención y Borrado de Datos

Este criterio evalúa si el proveedor cuenta con políticas claras y verificables sobre la retención, conservación y eliminación de datos personales.

El evaluador debe revisar si el proveedor:

- Define períodos de retención alineados con la finalidad del tratamiento.
- Se compromete contractualmente a borrar o devolver los datos al finalizar el contrato o cuando la entidad lo solicite.
- Establece procedimientos para la eliminación segura de la información.

Una política clara de retención y borrado es fundamental para evitar acumulación indebida de datos y riesgos de uso no autorizado.

Resultado de la Sección 3

La calificación de esta sección permite determinar si el proveedor ofrece garantías suficientes para proteger los datos personales de la ciudadanía, cumplir con la normativa colombiana y apoyar a la entidad distrital en el ejercicio de sus responsabilidades como responsable del tratamiento.

Una calificación baja indica riesgos legales y sancionatorios significativos, que deben ser mitigados mediante condiciones contractuales estrictas o que pueden llevar a la no recomendación del proveedor.

Forma 51. Privacidad y Protección de datos – Checklist Proveedores

SECCIÓN 3: PRIVACIDAD Y PROTECCIÓN DE DATOS (Peso: 20%)

ID	Criterio de Evaluación	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia/Referencia
3.1	**Claridad sobre Titularidad de Datos.**	No hay claridad contractual sobre la propiedad de los datos (entrenamiento, operación, metadatos).	La titularidad está definida, pero con ambigüedades sobre metadatos o datos derivados.	Contrato define claramente la titularidad de todos los datos y metadatos generados.	0	Documento o referencia
3.2	**Data Processing Agreement (DPA).**	No ofrece un DPA o se niega a suscribirlo.	Ofrece un DPA básico que requiere ajustes significativos para cumplir con la Ley 1581.	Ofrece un DPA robusto, alineado con la Ley 1581, que establece roles, obligaciones y mecanismos de auditoría.	0	Documento o referencia
3.3	**Gestión de Derechos de los Titulares (Habeas Data).**	No tiene procedimientos para atender derechos de acceso, rectificación, supresión y oposición.	Tiene procedimientos, pero son manuales, lentos o no garantizan los plazos legales (15 días).	Tiene procedimientos operativos eficientes y canales claros para que la entidad y los ciudadanos ejerzan los derechos de habeas data.	0	Documento o referencia
3.4	**Políticas de Retención y Borrado de Datos.**	No tiene políticas definidas de retención y borrado.	Tiene políticas genéricas no específicas al proyecto de IA.	Tiene políticas específicas de retención y borrado alineadas con la finalidad del tratamiento y se compromete.	0	Documento o referencia

CALIFICACIÓN SECCIÓN 3: 0/8

4. Seguridad y Robustez

Esta sección del checklist evalúa si el proveedor de soluciones de inteligencia artificial cuenta con condiciones mínimas de seguridad de la información y robustez técnica, que permitan proteger los datos, garantizar la continuidad del servicio y responder adecuadamente ante incidentes de seguridad.

En el contexto del Distrito Capital de Bogotá, esta sección busca asegurar que los sistemas de IA no introduzcan riesgos inaceptables para la operación institucional, la información pública o la confianza ciudadana, sin exigir niveles de madurez propios de entornos altamente regulados internacionales.

4.1 Certificaciones de Seguridad de la Información

Este criterio evalúa si el proveedor cuenta con certificaciones formales o evidencias equivalentes en materia de seguridad de la información.

El evaluador debe verificar si el proveedor:

- Cuenta con certificación ISO/IEC 27001 vigente o en proceso de implementación, o
- Dispone de controles documentados de seguridad de la información que cubran, al menos:
    - Gestión de accesos.
    - Protección de la confidencialidad, integridad y disponibilidad de la información.
    - Copias de respaldo y recuperación ante fallos.

La certificación no es obligatoria en todos los casos, pero la ausencia total de controles de seguridad documentados representa un riesgo significativo para la entidad distrital.

4.2 Pruebas de Robustez y Seguridad de IA

Este criterio evalúa si el proveedor ha realizado pruebas básicas para verificar la robustez técnica y la seguridad del sistema de IA frente a fallos, errores o comportamientos no esperados.

El evaluador debe revisar si el proveedor:

- Ha probado el comportamiento del sistema frente a datos incompletos, erróneos o fuera del escenario esperado.
- Ha identificado posibles fallos que puedan afectar la calidad del servicio o generar decisiones incorrectas.
- Puede documentar, aunque sea de forma básica, los resultados de estas pruebas y las acciones correctivas adoptadas.

No se exige la aplicación de pruebas avanzadas o especializadas, pero sí evidencia de que el proveedor comprende y gestiona los riesgos técnicos propios de sistemas de IA.

4.3 Respuesta a Incidentes de Seguridad

Este criterio evalúa si el proveedor cuenta con un protocolo claro para la gestión de incidentes de seguridad, incluyendo aquellos que involucren datos personales o interrupciones del servicio.

El evaluador debe verificar si el proveedor:

- Tiene definido un procedimiento para identificar, reportar y gestionar incidentes de seguridad.
- Se compromete a notificar oportunamente a la entidad distrital ante incidentes que afecten datos o la operación del sistema.
- Ofrece soporte para la contención, análisis y remediación del incidente.

Este criterio es clave para garantizar que la entidad pueda actuar de manera oportuna ante eventos que comprometan la información o la continuidad del servicio público.

Resultado de la Sección 4

La calificación de esta sección permite determinar si el proveedor ofrece un nivel aceptable de seguridad y robustez para operar sistemas de inteligencia artificial en el entorno institucional del Distrito Capital.

Una calificación baja indica riesgos relevantes en términos de seguridad de la información, interrupción del servicio y exposición a incidentes, lo que puede requerir condiciones contractuales adicionales o derivar en la no recomendación del proveedor.

Forma 52. Seguridad y Robustez – Checklist Proveedores

SECCIÓN 4: SEGURIDAD Y ROBUSTEZ (Peso: 20%)

ID	Criterio de Evaluación	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia/Referencia
4.1	**Certificaciones de Seguridad de la Información.**	No tiene certificaciones ni evidencias de prácticas de seguridad robustas.	Tiene certificaciones básicas o reportes de auditoría de seguridad con hallazgos menores no corregidos.	Tiene certificación ISO 27001 vigente o equivalente (ej.: SOC 2 Tipo II) y proporciona reportes recientes.	0	Documento o referencia
4.2	**Pruebas de Robustez y Seguridad de IA.**	No ha realizado pruebas de robustez o seguridad específicas para el modelo de IA.	Ha realizado pruebas básicas de rendimiento, pero no pruebas específicas para ataques adversarios o sesgos.	Ha realizado pruebas de robustez (resistencia a datos fuera de distribución) y seguridad (ej.: evaluación de vulnerabilidades a ataques adversarios) con resultados documentados.	0	Documento o referencia
4.3	**Respuesta a Incidentes de Seguridad.**	No tiene un protocolo documentado de respuesta a incidentes.	Tiene un protocolo básico, pero no está actualizado o no se ha probado.	Tiene un protocolo completo, actualizado y probado regularmente.	0	Documento o referencia

CALIFICACIÓN SECCIÓN 4: 0/6

5. Auditoría y Rendición de Cuentas

Esta sección del checklist evalúa si el proveedor de soluciones de inteligencia artificial garantiza condiciones mínimas de auditabilidad y trazabilidad, indispensables para que las entidades del Distrito Capital puedan ejercer control, supervisión y rendición de cuentas sobre los sistemas de IA utilizados en la prestación de servicios públicos.

En el sector público, la imposibilidad de auditar o explicar el funcionamiento de un sistema algorítmico representa un riesgo institucional crítico, tanto desde el punto de vista legal como desde la perspectiva de la confianza ciudadana.

5.1 Derecho a Auditoría

Este criterio evalúa si el proveedor acepta y facilita el derecho de la entidad distrital a auditar el sistema de IA, directamente o a través de terceros autorizados.

El evaluador debe verificar si el proveedor:

- Acepta contractualmente cláusulas de derecho a auditoría por parte de la entidad o de organismos de control.
- Se compromete a facilitar información, documentación y evidencias necesarias para procesos de auditoría técnica, jurídica o administrativa.
- Coopera con auditorías realizadas de forma remota o presencial, cuando así se requiera.

Este criterio no implica acceso irrestricto a información sensible del proveedor, pero sí garantiza que la entidad pueda verificar el cumplimiento de obligaciones contractuales, legales y éticas.

5.2 Trazabilidad de Decisiones

Este criterio evalúa si el sistema de inteligencia artificial permite reconstruir y explicar las decisiones o resultados generados, especialmente cuando estos tienen impacto sobre la ciudadanía o los procesos administrativos.

El evaluador debe revisar si el proveedor:

- Registra información básica sobre las entradas, salidas y resultados generados por el sistema.
- Permite identificar la versión del modelo utilizada en cada período de operación.
- Conserva registros (logs) que permitan analizar decisiones relevantes ante requerimientos internos o externos.

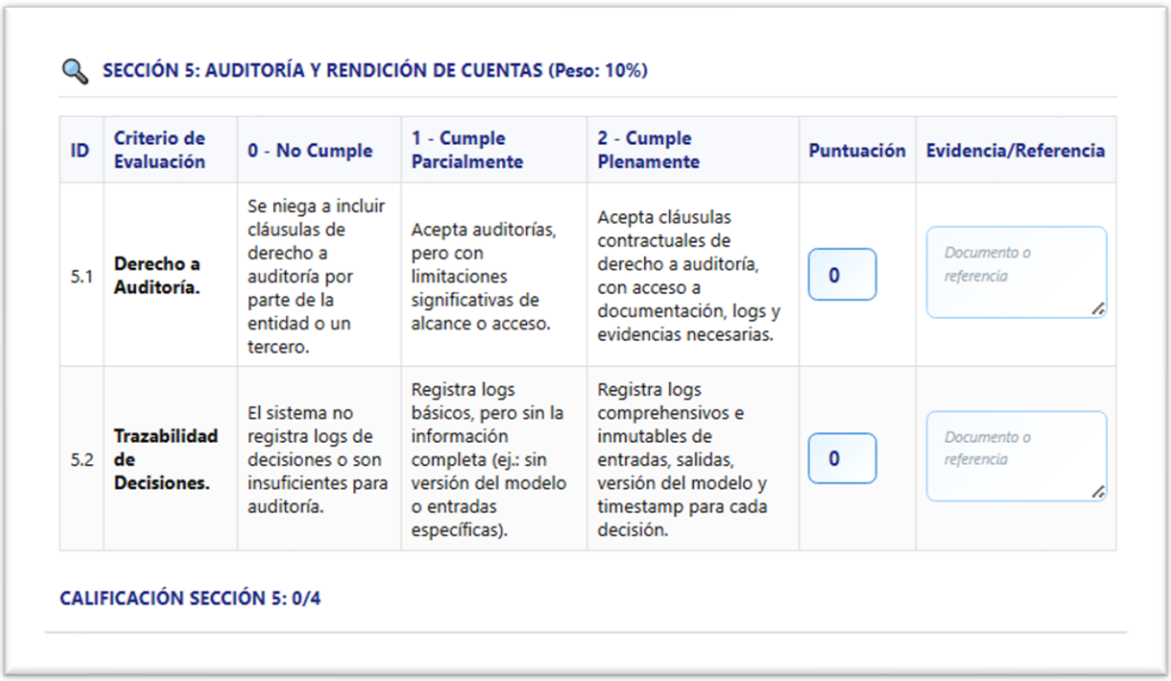
La trazabilidad es fundamental para garantizar explicabilidad, control posterior y rendición de cuentas, especialmente frente a reclamaciones ciudadanas, acciones judiciales o requerimientos de entes de control.

Resultado de la Sección 5

La calificación de esta sección permite determinar si el proveedor ofrece garantías suficientes de transparencia operativa y control institucional, acordes con los principios de publicidad, responsabilidad y control propios de la administración pública.

Una calificación baja indica riesgos significativos en términos de opacidad algorítmica, imposibilidad de auditoría y dificultad para responder ante reclamaciones o investigaciones, lo que puede derivar en condiciones contractuales estrictas o en la no recomendación del proveedor.

Forma 53. Auditoría y Rendición de cuentas – Checklist Proveedores



SECCIÓN 5: AUDITORÍA Y RENDICIÓN DE CUENTAS (Peso: 10%)

ID	Criterio de Evaluación	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia/Referencia
5.1	**Derecho a Auditoría.**	Se niega a incluir cláusulas de derecho a auditoría por parte de la entidad o un tercero.	Acepta auditorías, pero con limitaciones significativas de alcance o acceso.	Acepta cláusulas contractuales de derecho a auditoría, con acceso a documentación, logs y evidencias necesarias.	0	Documento o referencia
5.2	**Trazabilidad de Decisiones.**	El sistema no registra logs de decisiones o son insuficientes para auditoría.	Registra logs básicos, pero sin la información completa (ej.: sin versión del modelo o entradas específicas).	Registra logs comprehensivos e inmutables de entradas, salidas, versión del modelo y timestamp para cada decisión.	0	Documento o referencia

CALIFICACIÓN SECCIÓN 5: 0/4

6. Calidad del Servicio y Soporte

Esta sección del checklist evalúa si el proveedor de soluciones de inteligencia artificial cuenta con capacidades operativas suficientes para garantizar la continuidad, sostenibilidad y adecuada apropiación institucional del sistema durante todo su ciclo de vida.

En el contexto de las entidades públicas del Distrito Capital de Bogotá, la calidad del servicio y el soporte no solo inciden en la operación técnica, sino también en la reducción de dependencias del proveedor, la gestión del cambio organizacional y la sostenibilidad del proyecto en el tiempo.

6.1 Acuerdos de Nivel de Servicio (SLA)

Este criterio evalúa si el proveedor ofrece Acuerdos de Nivel de Servicio (Service Level Agreements – SLA) claros, medibles y exigibles contractualmente.

El evaluador debe verificar si el proveedor:

- Define compromisos explícitos de disponibilidad del servicio (por ejemplo, porcentaje de tiempo en operación).
- Establece tiempos de respuesta y resolución de incidentes, diferenciados por niveles de severidad.
- Incluye mecanismos de seguimiento, reporte y, cuando aplique, consecuencias por incumplimiento.

La existencia de SLA claros permite a la entidad distrital gestionar la operación del sistema con criterios objetivos y reducir riesgos de interrupción del servicio público.

6.2 Soporte Técnico y Transferencia de Conocimiento

Este criterio evalúa la capacidad del proveedor para acompañar técnica y operativamente a la entidad durante la implementación y operación del sistema, así como para facilitar la transferencia de conocimiento.

El evaluador debe revisar si el proveedor:

- Ofrece soporte técnico estructurado, indicando canales de atención y horarios.
- Proporciona capacitación básica a usuarios, operadores o administradores del sistema.
- Entrega documentación suficiente que permita a la entidad comprender y operar la solución sin dependencia excesiva del proveedor.

Este criterio es clave para fortalecer la autonomía institucional y garantizar la sostenibilidad del uso de la IA en el entorno distrital.

6.3 Gestión de Cambios y Roadmap

Este criterio evalúa si el proveedor cuenta con una gestión estructurada de cambios y si comunica de manera transparente la evolución futura de la solución.

El evaluador debe verificar si el proveedor:

- Informa oportunamente sobre actualizaciones, cambios funcionales o técnicos del sistema.

- Dispone de un roadmap de evolución del producto, al menos a nivel general.
- Permite a la entidad evaluar el impacto de cambios relevantes antes de su implementación.

La gestión adecuada de cambios es fundamental para evitar afectaciones inesperadas en procesos administrativos, decisiones públicas o servicios a la ciudadanía.

Resultado de la Sección 6

La calificación de esta sección permite determinar si el proveedor ofrece condiciones adecuadas de soporte, continuidad y acompañamiento, acordes con las necesidades operativas de una entidad pública del Distrito Capital.

Una calificación baja indica riesgos en términos de dependencia del proveedor, falta de soporte efectivo y dificultades para sostener el sistema en el tiempo, lo que puede requerir ajustes contractuales o llevar a la no recomendación del proveedor.

Forma 54. Calidad del servicio y soporte – Checklist Proveedores

SECCIÓN 6: CALIDAD DEL SERVICIO Y SOPORTE (Peso: 15%)

ID	Criterio de Evaluación	0 - No Cumple	1 - Cumple Parcialmente	2 - Cumple Plenamente	Puntuación	Evidencia/Referencia
6.1	**Acuerdos de Nivel de Servicio (SLA).**	No ofrece SLA o los compromisos no son cuantificables.	Ofrece SLA con métricas básicas, pero sin penalizaciones claras por incumplimiento.	Ofrece SLA robustos con métricas clave (uptime, tiempo de respuesta) y compensaciones definidas por incumplimiento.	0	Documento o referencia
6.2	**Soporte Técnico y Transferencia de Conocimiento.**	No ofrece un plan de soporte adecuado o canales limitados.	Ofrece soporte en horario laboral estándar y documentación básica.	Ofrece soporte prioritario, múltiples canales y un plan de transferencia de conocimiento (capacitación, documentación técnica).	0	Documento o referencia
6.3	**Gestión de Cambios y Roadmap.**	No comparte roadmap o notifica cambios con poca antelación.	Comparte un roadmap de alto nivel y notifica cambios con antelación moderada.	Comparte un roadmap detallado, notifica cambios con >30 días de antelación y permite la evaluación de impacto por parte de la entidad.	0	Documento o referencia

CALIFICACIÓN SECCIÓN 6: 0/6

Informe Final de Evaluación

El presente Informe Final de Evaluación consolida los resultados obtenidos a partir de la aplicación del Checklist de Evaluación de Proveedores de Inteligencia Artificial, en el marco del Framework de Gobernanza de IA del Distrito Capital de Bogotá.

Su propósito es apoyar la toma de decisiones informada sobre la viabilidad técnica, legal, ética y operativa del proveedor evaluado, así como documentar de manera trazable y transparente los criterios considerados, las fortalezas identificadas y los riesgos asociados a la eventual contratación o adopción de la solución de IA.

Este informe constituye un insumo formal para procesos contractuales, decisiones directivas y eventuales auditorías, y puede ser utilizado como respaldo ante entes de control.

Verificación de Criterios Descalificatorios

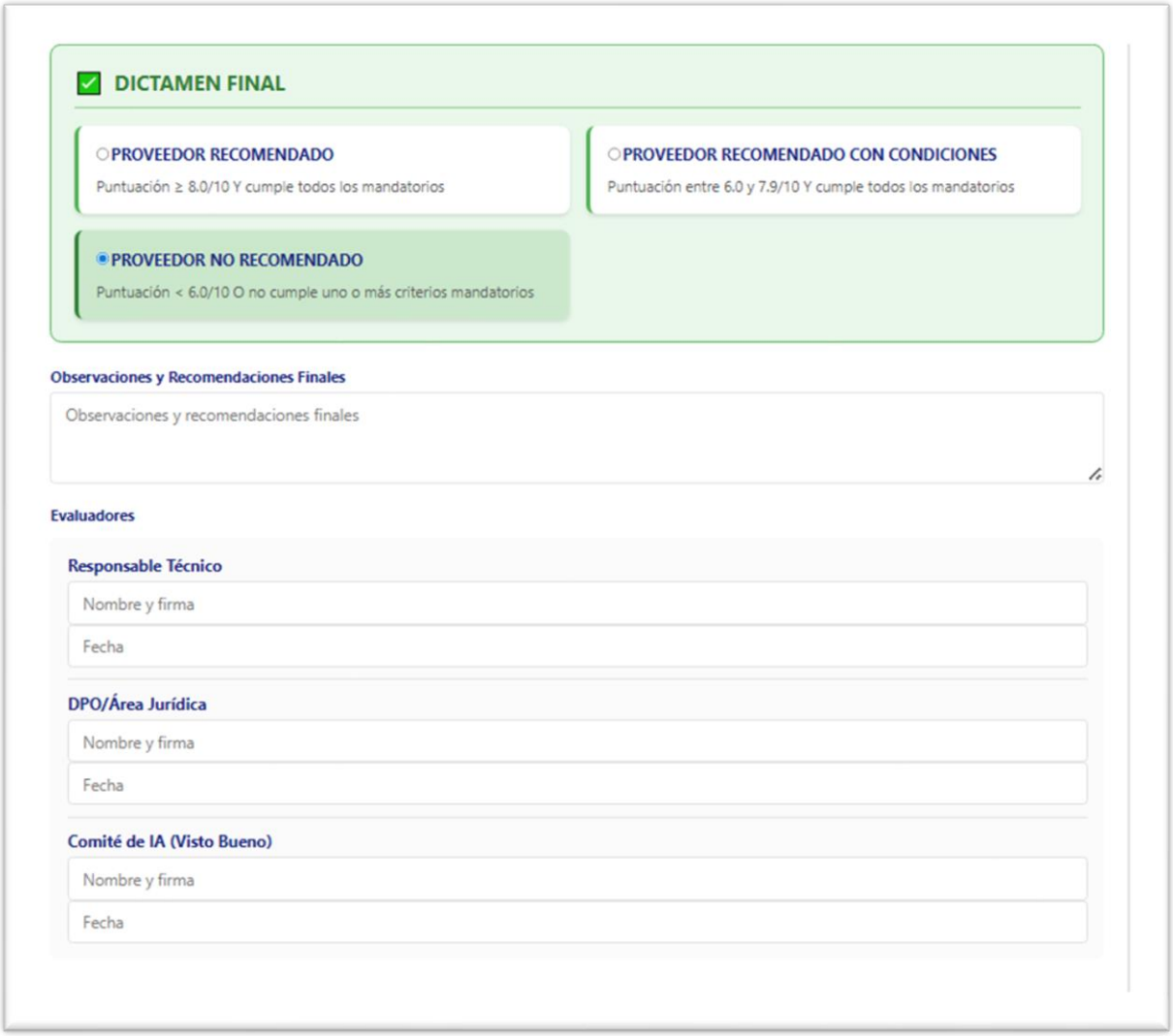
La verificación de criterios descalificatorios tiene como objetivo establecer un umbral mínimo obligatorio de cumplimiento para los proveedores de soluciones de inteligencia artificial que aspiren a contratar o prestar servicios a entidades del Distrito Capital de Bogotá.

Estos criterios representan requisitos habilitantes desde el punto de vista legal, operativo y de control institucional.

El incumplimiento de cualquiera de ellos implica la descalificación inmediata del proveedor, sin que sea procedente continuar con la evaluación ponderada del checklist.

Este enfoque es coherente con los principios de legalidad, responsabilidad administrativa y gestión del riesgo público.

Forma 55. Evaluación Final – Checklist Proveedores



Anexo F6. Manual Funcional Del Toolkit Model Cards

Introducción

El Model Card es una herramienta de documentación estandarizada que permite describir, de forma clara y comprensible, los modelos de inteligencia artificial utilizados por las entidades del Distrito Capital, incluyendo su propósito, funcionamiento general, datos utilizados, riesgos, controles y condiciones de uso.

Su objetivo principal es garantizar transparencia, trazabilidad, supervisión institucional y rendición de cuentas, asegurando que los modelos de IA se utilicen de manera responsable, ética y alineada con el interés público.

Dentro del Framework de Gobernanza de IA, el Model Card actúa como el documento oficial de referencia del modelo, acompañándolo durante todo su ciclo de vida y sirviendo como insumo para auditorías, decisiones directivas y comunicación con la ciudadanía.

1. Información General

Objetivo: La sección Información General tiene como propósito identificar de manera única el modelo de inteligencia artificial, establecer su responsabilidad institucional, y ubicarlo claramente dentro de su ciclo de vida.

Esta información es clave para garantizar trazabilidad, rendición de cuentas y control institucional sobre el uso del modelo.

Campo: Nombre del modelo

El nombre oficial con el que se identifica el modelo de IA dentro de la entidad.

Cómo diligenciarlo correctamente:

- Debe ser claro, descriptivo y comprensible.
- Debe reflejar la función principal del modelo.

- Evitar siglas internas poco comprensibles para terceros.

Campo: Versión

La versión actual del modelo, que permita identificar cambios y evoluciones en el tiempo.

- Utilizar un esquema sencillo de versionamiento (v1.0, v1.1, v2.0).
- Actualizar este campo cada vez que se realicen cambios relevantes en el modelo.

Campo: Entidad responsable

El nombre de la entidad del Distrito Capital que es responsable del modelo.

- Debe corresponder a la entidad que responde legal, técnica y administrativamente por el modelo.
- No debe confundirse con proveedores o terceros.

Campo: Área usuaria

El área, dependencia o dirección que utiliza directamente el modelo en su operación diaria.

- Identificar el área que toma decisiones o ejecuta procesos apoyados por el modelo.
- Puede ser distinta del área técnica o de TI.

Campo: Fecha de creación

La fecha en la que el modelo fue creado, adoptado o puesto en funcionamiento por primera vez.

- Usar el formato dd/mm/aaaa.
- Corresponde al inicio del ciclo de vida del modelo.

Campo: Fecha de actualización

La fecha más reciente en la que el modelo o su documentación fue actualizada.

- Usar el formato dd/mm/aaaa.
- Actualizar este campo cada vez que se realicen ajustes técnicos, funcionales o documentales.

Campo: Estado del ciclo de vida

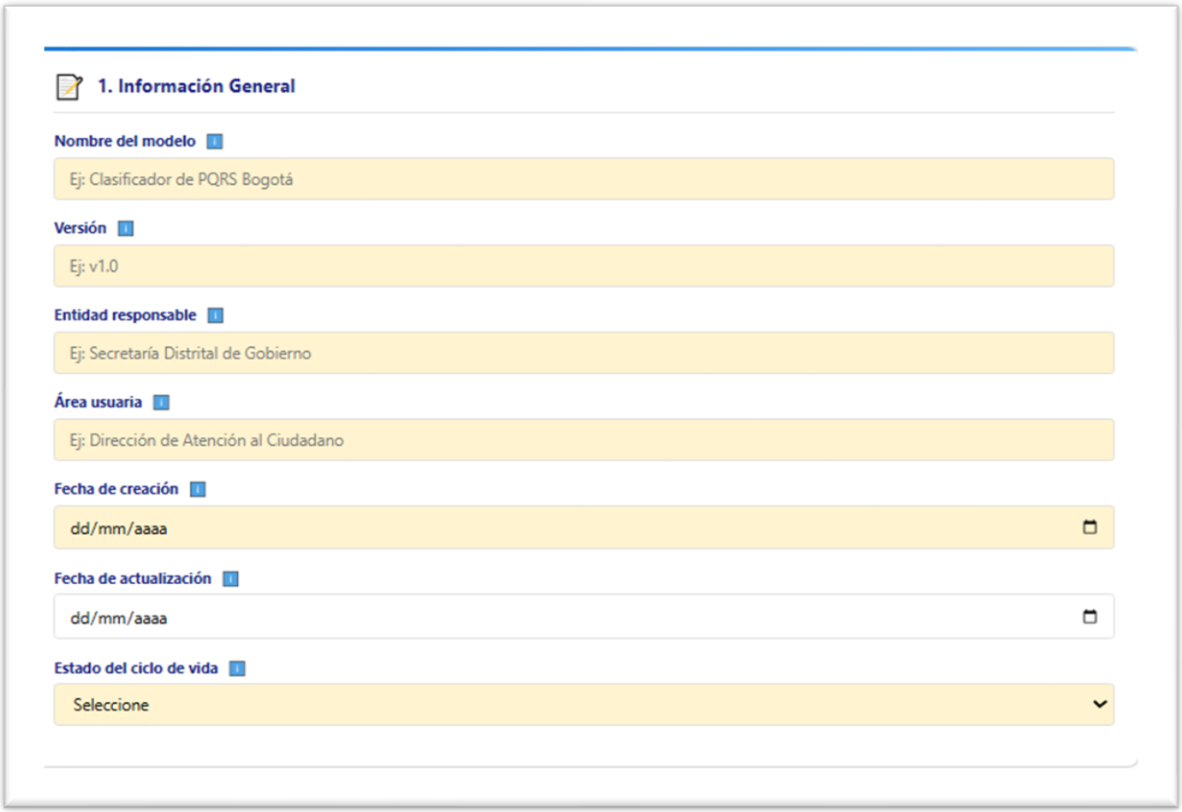
La etapa actual en la que se encuentra el modelo dentro de su ciclo de vida institucional.

Seleccionar el estado que mejor represente la situación actual del modelo, por ejemplo:

- En diseño
- En pruebas
- En producción
- En monitoreo
- En proceso de retiro

Este campo permite a la entidad controlar qué modelos están activos, en evaluación o próximos a su desactivación.

Forma 56. Sección 1 – Información General – Model Cards



2. Propósito del Modelo

Objetivo: La sección Propósito del Modelo tiene como finalidad explicar de manera clara y comprensible para qué existe el modelo de inteligencia artificial, qué problema busca resolver y en qué contexto institucional será utilizado.

Definir correctamente el propósito evita usos indebidos, expectativas irreales y riesgos asociados a la aplicación del modelo fuera de su contexto autorizado.

Campo: Objetivo del modelo

Una descripción clara del objetivo principal que persigue el modelo de IA.

Cómo diligenciarlo correctamente:

- Explicar qué busca lograr el modelo, no cómo lo hace técnicamente.
- Redactar en lenguaje sencillo, comprensible para personas no técnicas.
- Enfocarse en el valor que aporta al servicio público o a la gestión institucional.

Campo: Caso de uso

La descripción del caso de uso específico en el cual se empleará el modelo.

Cómo diligenciarlo correctamente:

- Indicar una situación concreta y real.
- Evitar descripciones genéricas o abstractas.
- Alinear este campo con el IA Use Case Canvas del framework.

Campo: Alcance

La delimitación clara de hasta dónde llega y hasta dónde NO llega el modelo.

Cómo diligenciarlo correctamente:

- Indicar qué decisiones apoya y cuáles no.
- Aclarar si el modelo recomienda, clasifica o prioriza, pero no decide de forma autónoma.
- Especificar límites funcionales y operativos.

Campo: Procesos institucionales impactados

La lista de procesos institucionales que se ven directa o indirectamente afectados por el uso del modelo.

Cómo diligenciarlo correctamente:

- Identificar procesos administrativos, misionales o de apoyo.
- Usar denominaciones institucionales reales.
- Este campo permite evaluar impactos organizacionales y riesgos operativos.

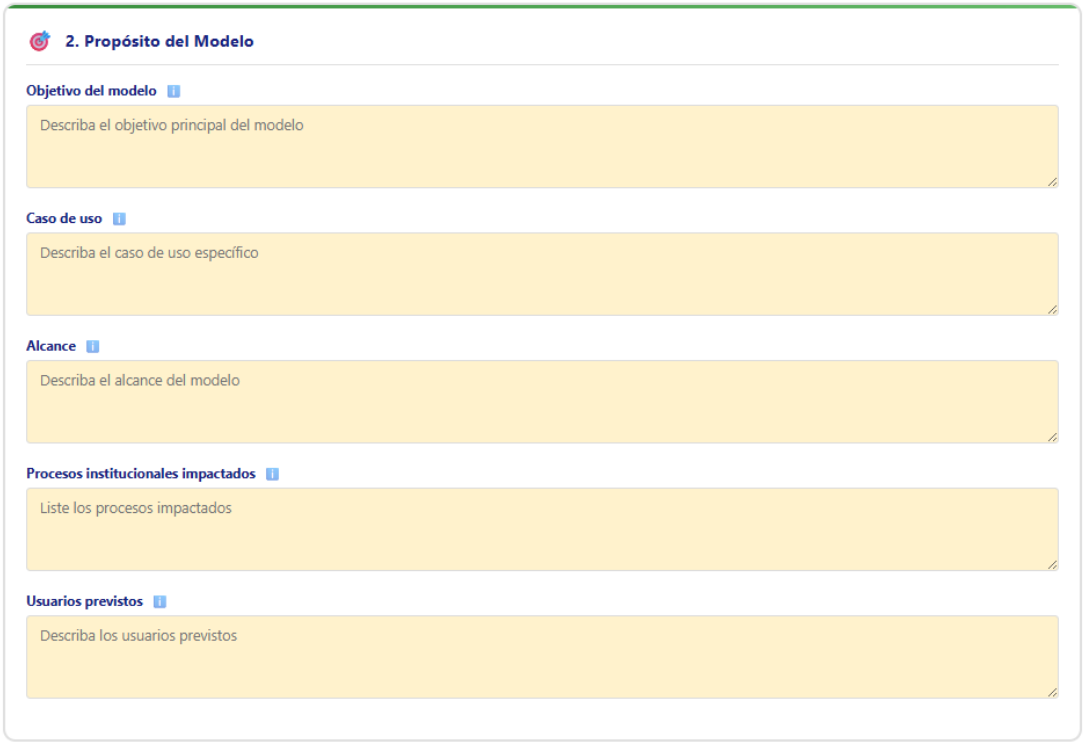
Campo: Usuarios previstos

La descripción de los usuarios que interactúan con el modelo o con sus resultados.

Cómo diligenciarlo correctamente:

- Diferenciar entre usuarios internos y externos.
- Describir perfiles generales (no nombres propios).
- Aclarar si los ciudadanos interactúan directa o indirectamente.

Forma 57. Sección 2 – Propósito del Modelo – Model Cards



3. Descripción Técnica

Objetivo: La sección Descripción Técnica tiene como finalidad documentar, de forma comprensible y transparente, las características técnicas esenciales del modelo de inteligencia artificial, sin requerir un nivel avanzado de conocimiento técnico por parte del lector.

Esta sección permite a la entidad:

- Comprender el tipo de tecnología utilizada.
- Facilitar procesos de auditoría técnica.
- Apoyar la gestión de riesgos y la toma de decisiones informadas.

- Garantizar transparencia sin comprometer secretos industriales.

Campo: Tipo de modelo

El tipo general de modelo de inteligencia artificial utilizado.

Cómo diligenciarlo correctamente:

- Seleccionar la opción que mejor describa el modelo según la lista disponible en la herramienta.
- No detallar algoritmos en este campo; solo la categoría general.
- Este campo permite clasificar el modelo dentro del ecosistema institucional de IA.

Campo: Algoritmo principal

El algoritmo o técnica principal sobre la cual se basa el modelo.

Cómo diligenciarlo correctamente:

- Indicar el algoritmo de forma clara y reconocible.
- No es necesario describir fórmulas matemáticas.
- Puede incluir técnicas modernas de aprendizaje automático o profundo.

Campo: Arquitectura

Una descripción general de la arquitectura del modelo, explicada en términos comprensibles.

Cómo diligenciarlo correctamente:

- Explicar la estructura general del modelo (capas, componentes, flujos).
- Usar lenguaje sencillo, evitando tecnicismos innecesarios.
- El objetivo es permitir comprensión, no replicación técnica.

Campo: Tecnologías utilizadas

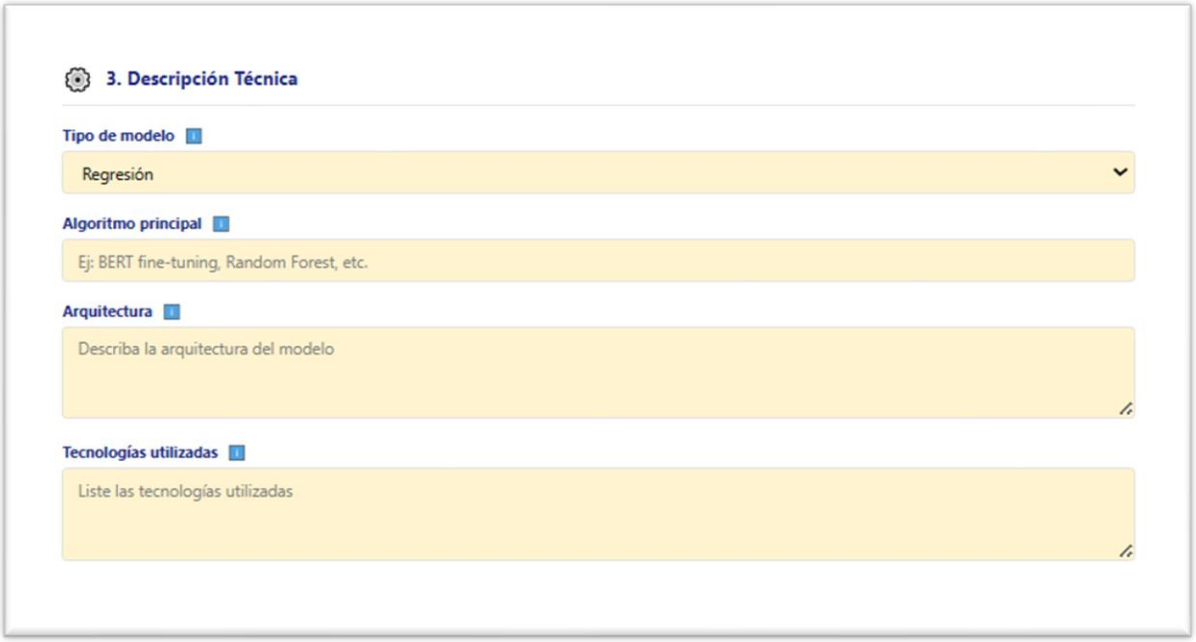
La lista de tecnologías, herramientas o plataformas empleadas para el desarrollo y operación del modelo.

Cómo diligenciarlo correctamente:

- Incluir lenguajes, frameworks y plataformas relevantes.
- No es necesario detallar versiones específicas, salvo que sea crítico.

- Este campo facilita evaluaciones de compatibilidad y sostenibilidad tecnológica.

Forma 58. Sección 3 – Descripción Técnica – Model Cards



4. Datos Utilizados

Objetivo: La sección Datos Utilizados tiene como propósito documentar de manera transparente los conjuntos de datos empleados en el desarrollo, validación y prueba del modelo de inteligencia artificial, así como su relación con la documentación formal de datos (Data Sheet).

Esta sección es fundamental para:

- Evaluar riesgos asociados a calidad, sesgos y representatividad de los datos.
- Garantizar cumplimiento de la normativa de protección de datos.
- Facilitar auditorías técnicas, legales y éticas.
- Fortalecer la confianza institucional y ciudadana en el uso del modelo.

Campo: Dataset de entrenamiento

Una descripción general del conjunto de datos utilizado para entrenar el modelo.

Cómo diligenciarlo correctamente:

- Indicar el tipo de datos utilizados (texto, registros administrativos, imágenes, etc.).
- Describir su origen de manera general (por ejemplo, datos institucionales, históricos).

- No incluir datos personales específicos ni información sensible detallada.

Campo: Dataset de validación

La descripción del conjunto de datos utilizado para validar el desempeño del modelo durante su desarrollo.

Cómo diligenciarlo correctamente:

- Indicar si el dataset es una partición del dataset de entrenamiento o un conjunto independiente.
- Explicar su finalidad: ajuste de parámetros y evaluación intermedia.
- Mantener un nivel descriptivo, no técnico.

Campo: Dataset de pruebas

La descripción del conjunto de datos utilizado para evaluar el desempeño final del modelo antes de su puesta en operación.

Cómo diligenciarlo correctamente:

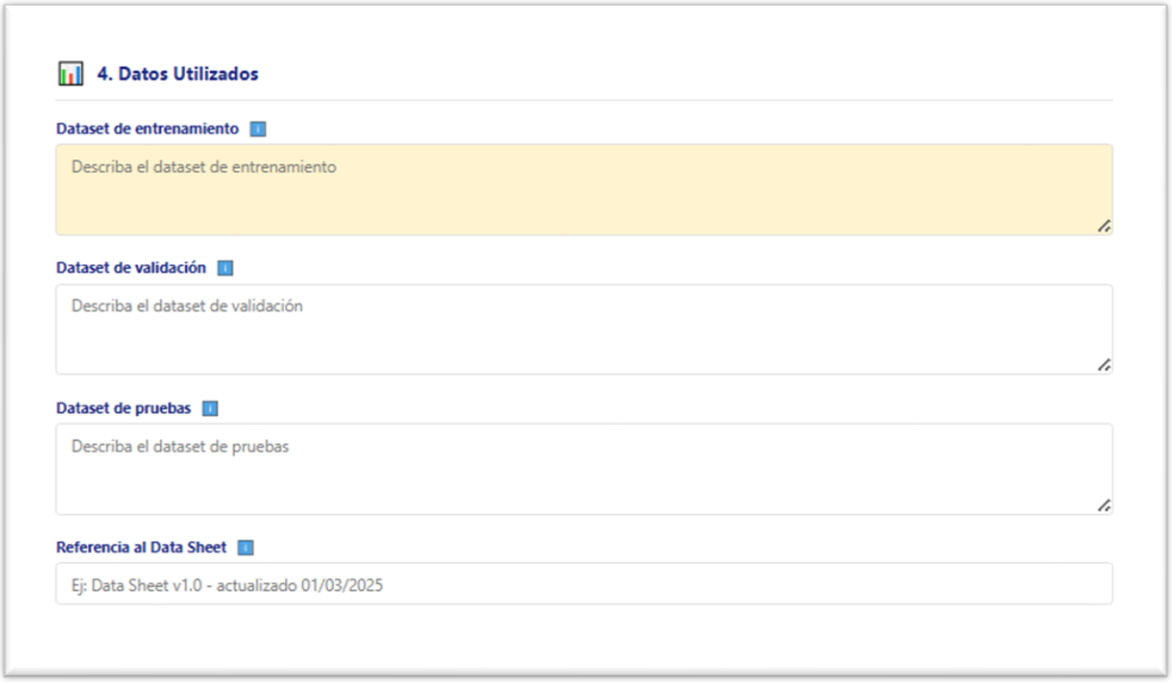
- Indicar que este dataset no fue utilizado durante el entrenamiento.
- Describir su función como evaluación final e independiente.
- Aclarar si representa escenarios reales de operación.

Campo: Referencia al Data Sheet

Cómo diligenciarlo correctamente:

- Indicar el nombre y versión del Data Sheet.
- Incluir la fecha de actualización.
- Este campo asegura coherencia entre herramientas del framework.

Forma 59. Sección 4 – Datos Utilizados – Model Cards



5. Métricas de Desempeño

Objetivo: La sección Métricas de Desempeño tiene como finalidad documentar de forma transparente y verificable el desempeño del modelo de inteligencia artificial, tanto a nivel general como en grupos específicos de población o contextos de uso.

Esta sección es clave para:

- Evaluar la efectividad real del modelo.
- Identificar posibles brechas de desempeño.
- Detectar riesgos de inequidad o sesgos.
- Sustentar decisiones de despliegue, ajuste o retiro del modelo.

5.1 Métricas Globales

En esta tabla se registran las métricas generales que resumen el desempeño global del modelo, calculadas sobre el conjunto de datos de prueba.

Cada fila representa una métrica distinta.

Columnas de la tabla

Métrica

El nombre de la métrica utilizada para evaluar el modelo.

Cómo diligenciarlo correctamente:

- Utilizar métricas estándar y reconocidas.
- Mantener consistencia entre versiones del modelo.
- Evitar métricas no explicables para usuarios no técnicos.

Ejemplos:

- Precisión
- Exactitud
- F1-score
- Recall

Valor

El valor numérico obtenido para la métrica.

Cómo diligenciarlo correctamente:

- Indicar el valor en formato decimal o porcentual.
- Usar el mismo formato en todas las métricas.
- Corresponder al dataset de pruebas.

Ejemplo:

0.89

Descripción

Una breve explicación, en lenguaje sencillo, de qué mide la métrica y por qué es relevante.

Cómo diligenciarlo correctamente:

- Explicar la métrica de forma comprensible para personas no técnicas.
- Relacionarla con el objetivo del modelo.

Ejemplo:

Mide la proporción de clasificaciones correctas realizadas por el modelo sobre el total de casos evaluados.

5.2 Métricas por Subpoblaciones

En esta tabla se registran métricas de desempeño desagregadas por subpoblaciones, con el fin de identificar posibles diferencias en el comportamiento del modelo entre distintos grupos o contextos.

El uso de esta tabla es fundamental para evaluar equidad y no discriminación.

Columnas de la tabla

Subpoblación

El grupo, contexto o segmento específico sobre el cual se calcula la métrica.

Cómo diligenciarlo correctamente:

- Definir subpoblaciones relevantes desde el punto de vista institucional.
- No utilizar datos sensibles identificables.
- Usar categorías agregadas.

Ejemplo:

Localidades rurales

Usuarios con alto volumen de solicitudes

Métrica

La métrica utilizada para evaluar el desempeño en esa subpoblación.

Cómo diligenciarlo correctamente:

- Puede ser la misma métrica usada a nivel global u otra más adecuada.
- Mantener claridad y consistencia.

Ejemplo:

F1-score

Valor

El valor obtenido para la métrica en la subpoblación definida.

Cómo diligenciarlo correctamente:

- Usar el mismo formato que en métricas globales.
- Asegurar que el cálculo sea técnicamente válido.

Ejemplo:

0.78

Observaciones

Comentarios cualitativos que ayuden a interpretar el resultado.

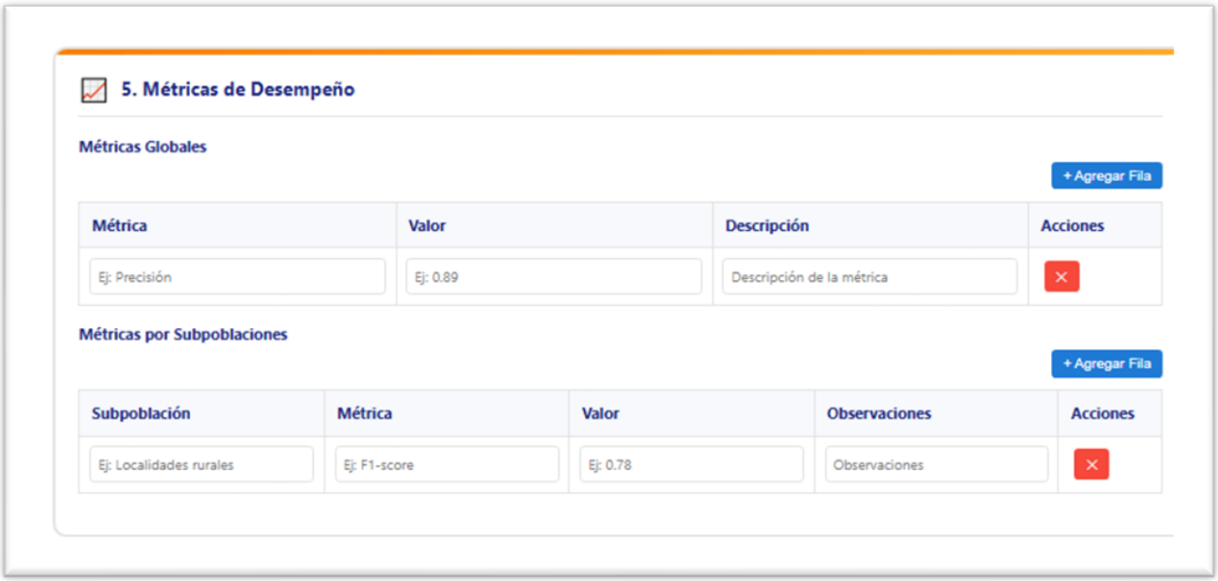
Cómo diligenciarlo correctamente:

- Indicar posibles causas de diferencias de desempeño.
- Señalar si se requiere seguimiento o ajustes.
- Usar lenguaje claro y no técnico.

Ejemplo:

Se observa menor desempeño debido a menor volumen de datos históricos en esta subpoblación.

Forma 60. Sección 5 – Métricas de Desempeño – Model Cards



6. Evaluación de Sesgos

Objetivo: La sección Evaluación de Sesgos tiene como propósito identificar, documentar y gestionar posibles sesgos presentes en el modelo de inteligencia artificial, con el fin de prevenir impactos negativos sobre la equidad, la no discriminación y los derechos de la ciudadanía.

Esta sección es fundamental para garantizar que el uso del modelo sea justo, proporcional y alineado con los principios de ética pública y derechos fundamentales.

Campo: Sesgos identificados

La descripción de los sesgos que hayan sido identificados durante el análisis del modelo.

Cómo diligenciarlo correctamente:

- Indicar los tipos de sesgos detectados (por ejemplo, sesgos por origen geográfico, frecuencia de datos, idioma).
- Reconocer explícitamente si no se identificaron sesgos relevantes.
- Evitar minimizar o ocultar limitaciones conocidas del modelo.

Ejemplo:

Se identificó un posible sesgo en la clasificación de solicitudes provenientes de zonas con menor volumen histórico de datos.

Campo: Resultados de pruebas de equidad

La descripción de los resultados obtenidos a partir de pruebas realizadas para evaluar la equidad del modelo.

Cómo diligenciarlo correctamente:

- Explicar de manera general cómo se evaluó la equidad (comparación entre grupos, análisis por subpoblaciones).
- Indicar si se observaron diferencias significativas en el desempeño.
- Utilizar lenguaje claro y comprensible.

Ejemplo:

Las pruebas de equidad muestran diferencias moderadas en el desempeño del modelo entre subpoblaciones, especialmente en contextos con menor representación de datos.

Campo: Medidas aplicadas para mitigar sesgos

Las acciones implementadas para reducir, mitigar o gestionar los sesgos identificados.

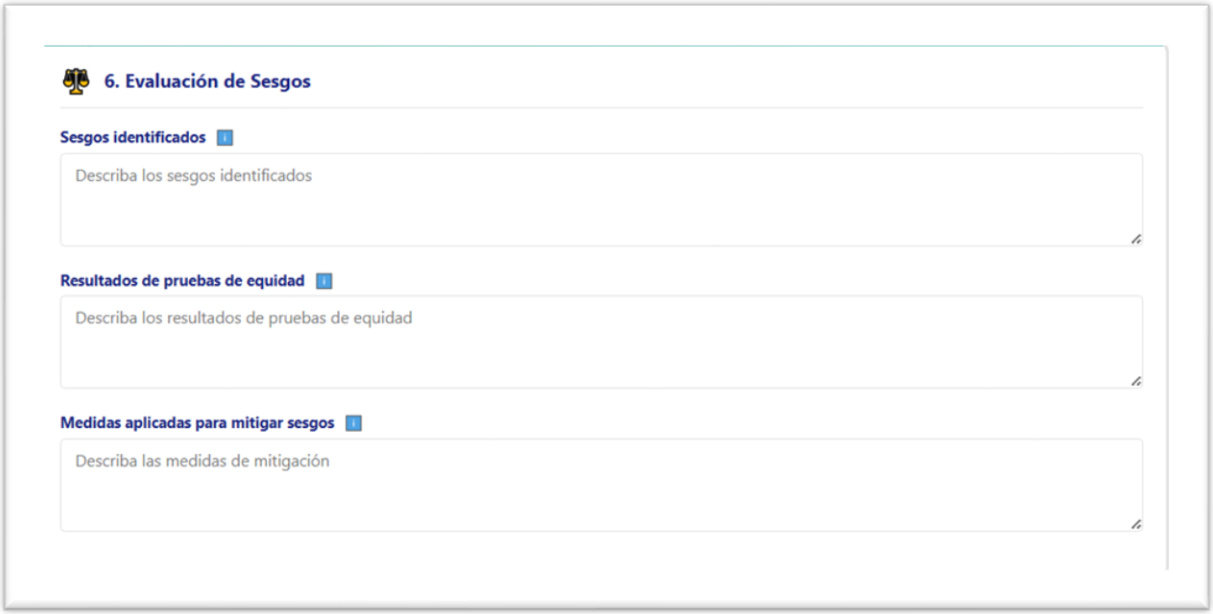
Cómo diligenciarlo correctamente:

- Describir medidas técnicas (ajustes de datos, recalibración del modelo).
- Incluir medidas organizacionales (supervisión humana, revisión periódica).
- Reconocer cuando los sesgos no pueden eliminarse completamente y explicar cómo se gestionan.

Ejemplo:

Se aplicaron técnicas de balanceo de datos y se estableció revisión humana para los casos clasificados con menor nivel de confianza.

Forma 61. Sección 6 – Evaluación de Sesgos – Model Cards



7. Riesgos del Modelo

Objetivo: Anticipar riesgos asociados al uso del modelo.

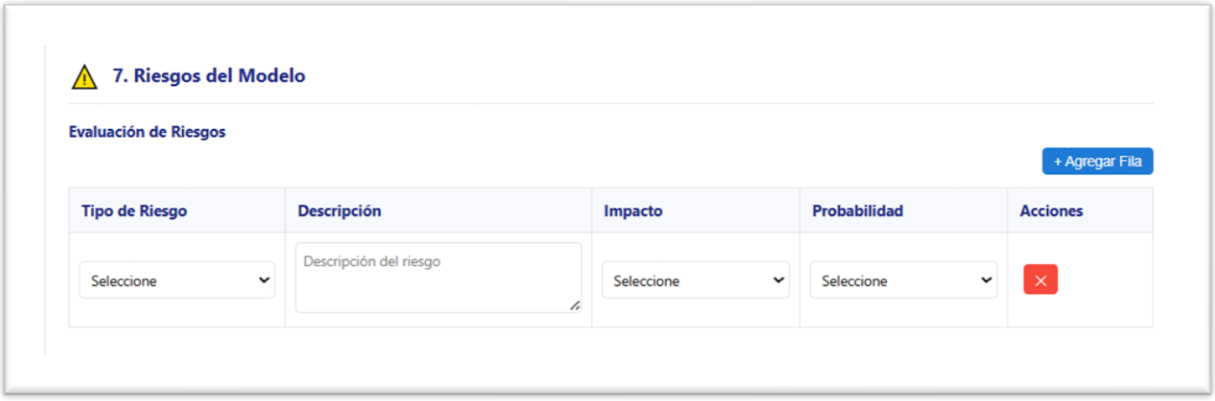
Identificar riesgos técnicos, legales, éticos u operativos relevantes.

Información a diligenciar:

- Riesgos identificados.
- Impactos potenciales.

- Probabilidad
- Acciones

Forma 62. Sección 7 – Riesgos del Modelo – Model Cards



8. Controles Implementados

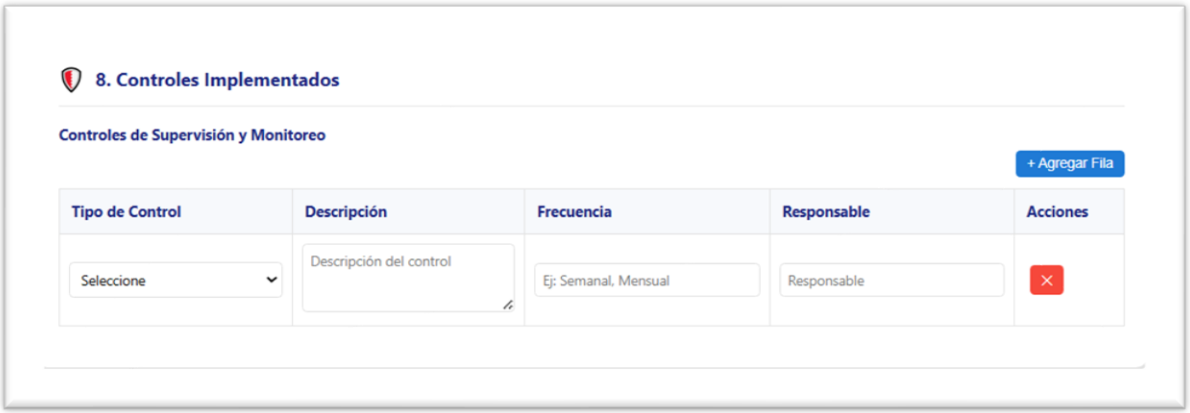
Objetivo: Mostrar cómo se gestionan los riesgos.

Describir las medidas técnicas y organizacionales implementadas para mitigar los riesgos.

Información a diligenciar:

- Tipo de control
- Descripción
- Frecuencia del control (mensual, semanal, diaria, etc)

Forma 63. Sección 8 – Controles Implementados – Model Cards



9. Explicabilidad y Transparencia

Objetivo: La sección Explicabilidad y Transparencia tiene como propósito garantizar que el funcionamiento del modelo de inteligencia artificial pueda ser comprendido, explicado y supervisado, especialmente cuando sus resultados influyen en decisiones administrativas o en la prestación de servicios públicos.

Esta sección refuerza los principios de publicidad, transparencia, rendición de cuentas y confianza ciudadana, propios de la administración pública.

Campo: Técnicas de explicabilidad utilizadas

Las técnicas empleadas para explicar cómo el modelo genera sus resultados o recomendaciones.

Cómo diligenciarlo correctamente:

- Indicar las técnicas utilizadas de manera general.
- No es necesario detallar implementaciones técnicas complejas.
- Aclarar si se utilizan técnicas automáticas o análisis interpretativos.

Ejemplos:

Análisis de características relevantes

Si no se utilizan técnicas específicas, debe indicarse claramente y explicar cómo se garantiza la comprensión del modelo por otros medios.

Campo: Información pública disponible

La descripción de la información sobre el modelo que es accesible al público o a los usuarios internos.

Cómo diligenciarlo correctamente:

- Indicar si existe documentación pública, guías, fichas informativas o avisos.
- Describir el tipo de información disponible (objetivo del modelo, uso permitido, limitaciones).
- No incluir información sensible o confidencial.

Ejemplo:

Se encuentra disponible una ficha informativa pública que describe el propósito del modelo, su alcance y los mecanismos de supervisión.

Campo: Mecanismos de rendición de cuentas

Los mecanismos institucionales que permiten responder por el uso, resultados y efectos del modelo.

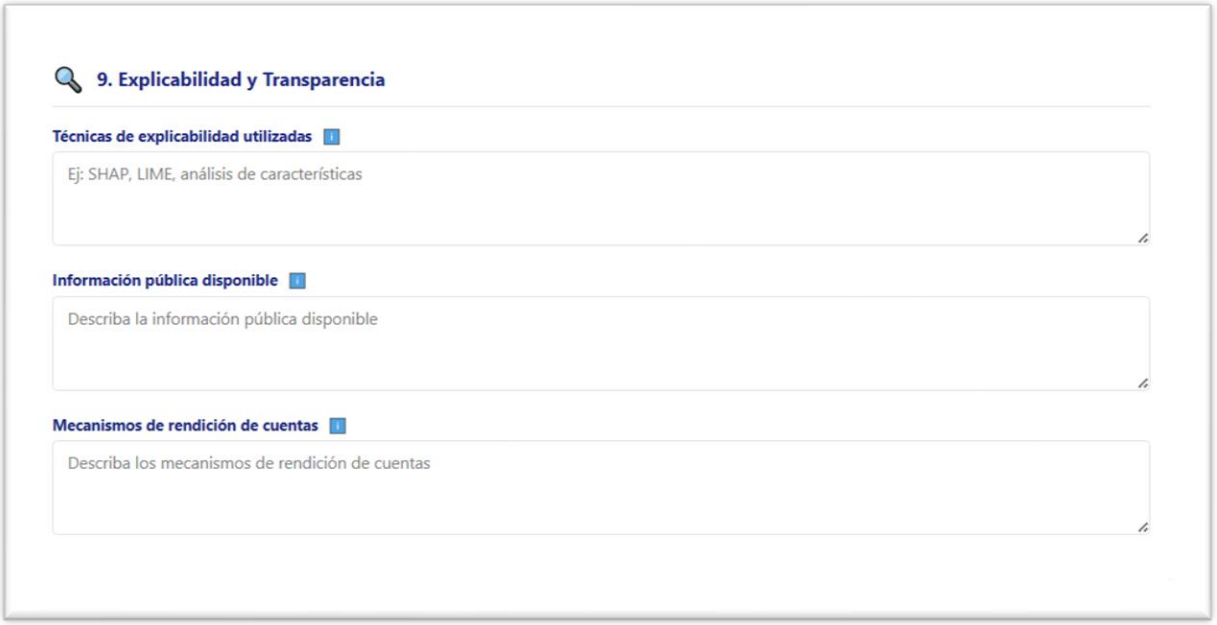
Cómo diligenciarlo correctamente:

- Describir canales de atención, revisión o reclamación.
- Indicar responsables institucionales.
- Explicar cómo se gestionan incidentes, errores o reclamos ciudadanos relacionados con el modelo.

Ejemplo:

El modelo cuenta con supervisión por parte del área usuaria y mecanismos de revisión administrativa ante reclamaciones ciudadanas.

Forma 64. Sección 9 – Explicabilidad y Transparencia – Model Cards



10. Reglas de Uso Responsable

Objetivo: Delimitar claramente usos permitidos y prohibidos.

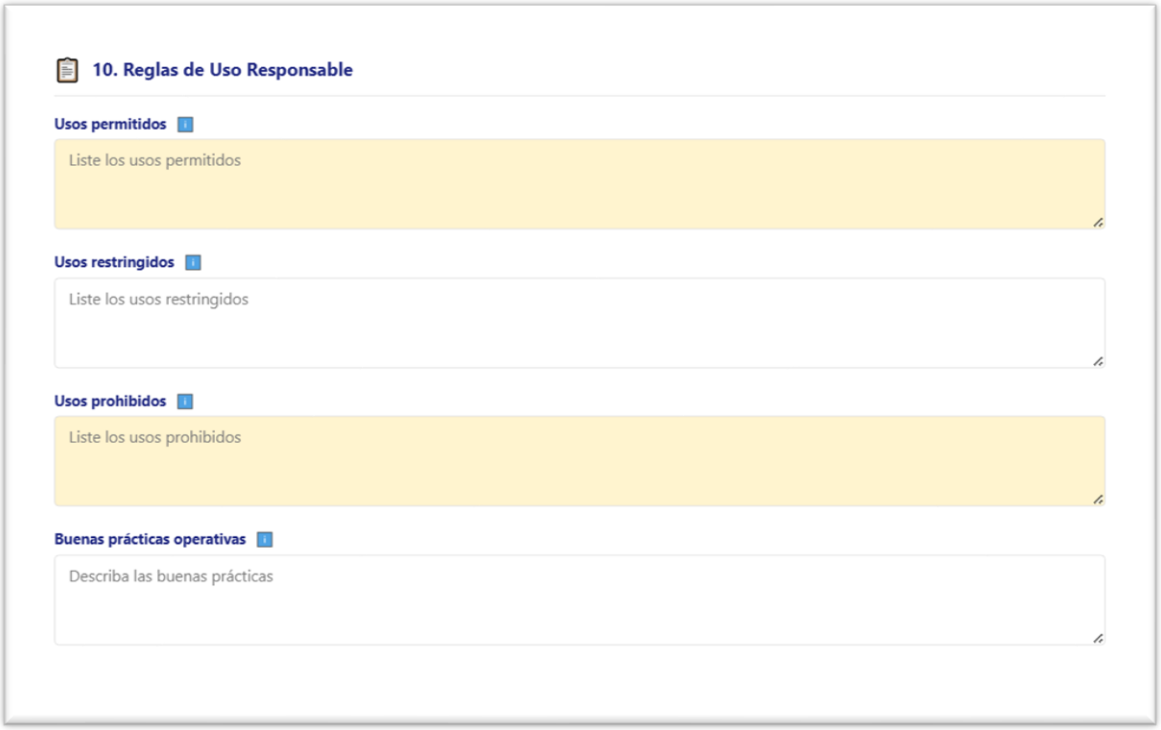
Definir explícitamente en qué casos puede y no puede usarse el modelo.

Información a diligenciar:

- Usos permitidos.
- Usos no permitidos.

- Restricciones específicas.

Forma 65. Sección 10 – Reglas de Uso Responsable – Model Cards



11. Entradas y Salidas del Modelo

Objetivo: Documentar el flujo de información.

Describir qué datos recibe el modelo y qué resultados genera.

Información a diligenciar:

- Tipos de entradas.
- Tipos de salidas.
- Uso de los resultados en decisiones humanas.

Forma 66. Sección 11 – Entradas y Salidas del Modelo – Model Cards

12. Monitoreo y Mantenimiento

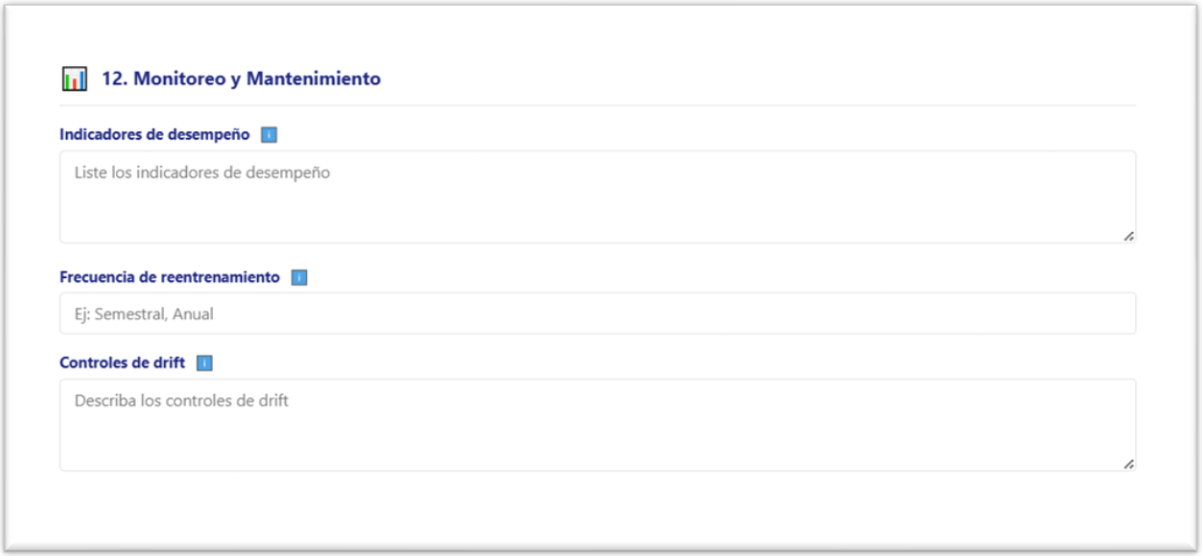
Objetivo: Asegurar la vigencia del modelo en el tiempo.

Definir cómo se revisa el desempeño del modelo y qué acciones se toman ante fallas.

Información a diligenciar:

- Frecuencia de revisión.
- Indicadores de alerta.
- Acciones correctivas.

Forma 67. Sección 12 – Monitoreo y Mantenimiento – Model Cards



13. Historial de Cambios

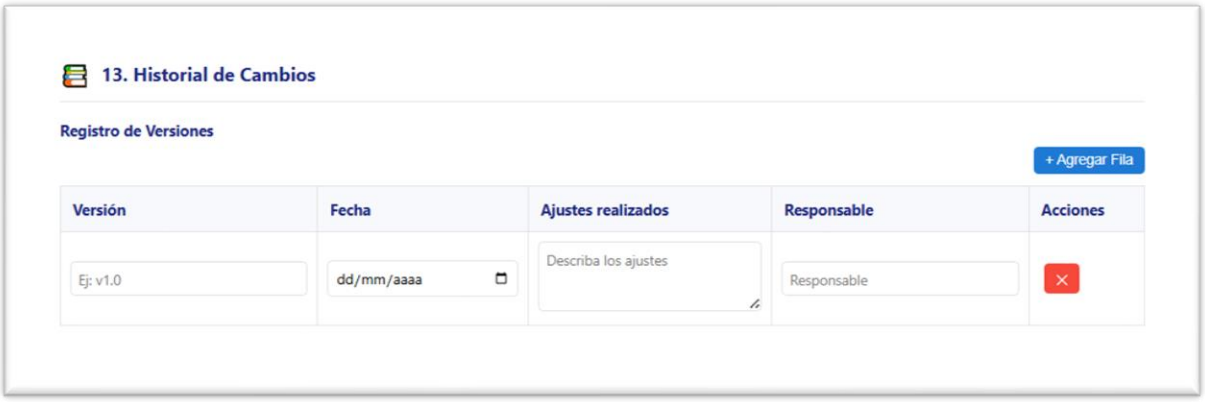
Objetivo: Registrar la evolución del modelo.

Documentar cambios relevantes realizados al modelo.

Información a diligenciar:

- Cambios realizados.
- Fecha.
- Justificación.

Forma 68. Sección 13 – Historial de Cambios – Model Cards



14. Aprobaciones

Objetivo: Formalizar la validación institucional.

Registrar la aprobación formal del uso del modelo.

Información a diligenciar:

- Responsable de aprobación.
- Fecha.
- Observaciones finales.

Forma 69. Sección 14 – Aprobaciones – Model Cards