ADMINISTRADORA DE LOS RECURSOS DEL SISTEMA GENERAL DE SEGURIDAD SOCIAL EN SALUD – ADRES

INFORME FINAL DE EJECUCIÓN - POC SISTEMA INTELIGENTE DE AUDITORÍA (SIA)

OBJETO CONTRACTUAL

Celebrar un convenio especial de cooperación que permita aunar esfuerzos para investigar, diseñar, desarrollar, implementar y brindar soporte a una solución tecnológica basada en inteligencia artificial, orientada a automatizar la revisión de cuentas médicas y gastos de servicios de salud en los procesos de auditoría a cargo de la ADRES; procurando la interoperabilidad con los sistemas de información necesarios, asegurando el cumplimiento de la normatividad vigente en materia de protección de datos personales y seguridad de la información, así como garantizando la transferencia y apropiación de conocimiento y nuevas tecnologías a favor de la Entidad para fortalecer de manera sostenible su capacidad de gestión institucional.

Bogotá D.C., octubre de 2025

Contenido

1. Resumen Ejecutivo 3
2. Introducción. 4
3. Aspectos generales de la prueba de concepto. 6
4. Proceso de convocatoria y vinculación de participantes 6
   4.1. Manifestación de interés y firma de acuerdos de entendimiento. 8
   4.2. Cronograma de la prueba de concepto. 9
5. Metodología de la prueba de concepto y estrategias de acompañamiento. 10
   5.1. Planteamiento metodológico en desarrollo de la PoC. 10
   5.2. Entregables esperados. 11
   5.3. Metodología de evaluación. 12
   5.4. Estrategia de acompañamiento por parte del equipo auditor de ADRES 16
   5.5. Estrategia de comunicaciones y notificaciones 17
6. Concepto de seguridad de la información emitido por el oficial de riesgo de ADRES 17
7. Desarrollo de la Prueba de Concepto 19
   7.1. Metodología Técnica y Desempeño de Infraestructura 19
   7.2. Historial de Participación en las Iteraciones 21
   7.3. Resultados preliminares durante las primeras cuatro iteraciones. 21
   7.4. Resultados de la Entrega final del Producto Mínimo Viable (PMV) 24
8. Conclusiones y Recomendaciones 28
   8.1. Conclusiones 28
   8.2. Recomendaciones 28

1. Resumen Ejecutivo

El proceso actual de auditoría de cuentas médicas en ADRES es predominantemente manual, lo que genera altos costos operativos, tiempos de respuesta prolongados y riesgo de errores humanos. En respuesta a este desafío y alineada con las políticas nacionales de transformación digital, la entidad formuló una Prueba de Concepto (PoC) para "analizar la viabilidad técnica, normativa y económica de implementar un prototipo de inteligencia artificial (IA)" que apoye la auditoría de reclamaciones por accidentes de tránsito no SOAT.

Se invitó a los cuatro principales proveedores de servicios en la nube (Google, Amazon Web Services, Oracle y Microsoft), quienes participaron a través de sus aliados locales en un proceso iterativo de seis fases para desarrollar un Producto Mínimo Viable (PMV).

El desarrollo de la PoC evidenció una evolución significativa. En las primeras iteraciones, todos los proveedores presentaron fallas críticas comunes, enfocándose en glosas administrativas con baja especificidad y una coincidencia casi nula (0%) con los códigos de glosa médica requeridos. Gracias a una estrategia de acompañamiento intensivo por parte del equipo auditor de ADRES, que incluyó capacitaciones y retroalimentación personalizada, los prototipos mejoraron su enfoque hacia la pertinencia médica.

La evaluación final del PMV se realizó sobre una muestra de 1.106 reclamaciones y se midió usando tres criterios ponderados: Precisión (50%), Desempeño (15%) y Eficiencia (35%).

El proveedor AWS (aliado Blend 360) demostró una madurez y un desempeño claramente superiores, alcanzando una calificación final de 99,8 sobre 100. AWS obtuvo los mejores resultados en todas las métricas clave: la mayor cobertura de procesamiento (96,7% de los ítems), la precisión más alta (F-Beta de 30,1), un desempeño sobresaliente en la justificación de glosas y la mayor eficiencia económica ajustada por su capacidad operativa real.

Google (aliado Servinformación) obtuvo el segundo lugar con 35,7 puntos, mostrando una buena capacidad de justificación, pero limitado por una menor cobertura (56,3%) y una eficiencia significativamente inferior. Microsoft (19,1 puntos) y Oracle (3,4 puntos) no alcanzaron los niveles mínimos de cobertura ni precisión requeridos.

La PoC confirma la viabilidad de implementar sistemas de IA para la auditoría médica. Se recomienda avanzar con el proveedor AWS, dada su idoneidad y madurez técnica demostradas para un escalamiento a producción.

2. Introducción.

El derecho a la salud, reconocido en la Constitución Política de Colombia y desarrollado a través de la jurisprudencia, constituye un eje fundamental del bienestar social y del cumplimiento de los fines esenciales del Estado. Así, la Ley 100 de 1993 estructuró el Sistema General de Seguridad Social en Salud (SGSSS) con el propósito de garantizar la prestación oportuna, eficaz y de calidad de los servicios de salud, asegurando cobertura universal y sostenibilidad financiera. A su vez, la Corte Constitucional, mediante la Sentencia T-760 de 2008, reconoció la naturaleza fundamental y autónoma del derecho a la salud, precisando que su garantía efectiva requiere acciones estatales coordinadas, recursos tecnológicos e institucionales adecuados, y una gestión eficiente de los recursos públicos destinados a su materialización.

En este marco, la Ley Estatutaria 1751 de 2015 consolidó el derecho fundamental a la salud y dispuso que su garantía debe sustentarse en una prestación integral que incluya promoción, prevención, atención, rehabilitación y paliación. Para materializar este mandato, se creó la la **Administradora de los Recursos del Sistema General de Seguridad Social en Salud (ADRES)** como una entidad de gestión que asegura el uso transparente, estratégico y racional de los recursos del sistema.

Creada por el artículo 66 de la Ley 1753 de 2015, la ADRES es una entidad de naturaleza especial, con autonomía administrativa y financiera, encargada de administrar y proteger los recursos del SGSSS. Desde su entrada en operación en 2017, la ADRES ha enfrentado el reto de gestionar un sistema de salud cada vez más complejo, en el que los procesos de auditoría médica especializada y validación de cuentas representan un componente crítico para la sostenibilidad financiera y la transparencia institucional.

Actualmente, el proceso de auditoría de cuentas médicas —que busca verificar la justificación de la pertinencia médica, oportunidad y legalidad de los pagos— y otros gastos en salud a cargo de la ADRES presenta limitaciones debido a su carácter predominantemente manual. La revisión individual de expedientes, la multiplicidad de formatos, la variabilidad en la calidad de los datos, la sensibilidad en el tratamiento de los datos clínicos y la alta carga documental, generan cuellos de botella que incrementan los costos operativos, prolongan los tiempos de respuesta y exponen el proceso a errores humanos. Frente a este escenario, la ADRES ha identificado una **oportunidad de innovación pública** orientada a la adopción de tecnologías de inteligencia artificial (IA) que apoyen la toma de decisiones y optimicen la gestión de auditorías médicas.

La **innovación pública digital**, como lo establece el Decreto 767 de 2022 y el Plan Estratégico 2023-2026 de la ADRES, constituye una herramienta fundamental para modernizar la

entidad, fortalecer la transparencia, mejorar la trazabilidad del gasto público y promover la eficiencia institucional. En coherencia con la Política de Gobierno Digital, la ADRES ha proyectado la incorporación de metodologías y soluciones tecnológicas que permitan automatizar procesos, realizar análisis predictivos y aprovechar los datos disponibles para orientar decisiones basadas en evidencia.

En este contexto, el uso de sistemas de inteligencia artificial ofrece un alto potencial transformador. Las tecnologías de arquitectura híbrida —que integran modelos de lenguaje a gran escala (LLM), mecanismos de recuperación aumentada de información (RAG), validaciones estructurales y reglas normativas codificadas— permiten procesar grandes volúmenes de datos y generar resultados explicables, auditables y alineados con los estándares vigentes en materia de auditoría en salud. Su implementación puede fortalecer la independencia tecnológica de la ADRES, agilizar los procesos de revisión, mejorar la precisión de las decisiones y racionalizar el gasto público en salud.

Esta iniciativa se enmarca en las políticas nacionales que promueven la transformación digital del Estado. En particular, el **Documento CONPES 3975 de 2019** formuló la Política Nacional para la Transformación Digital e Inteligencia Artificial, con el objetivo de generar valor público mediante el uso estratégico de tecnologías digitales en el sector público. De manera complementaria, el **CONPES 4144 de 2025** adoptó la Política Nacional de Inteligencia Artificial, orientada a fomentar proyectos piloto, fortalecer capacidades institucionales y evaluar los impactos de la IA, especialmente en sectores sensibles como el de la salud.

El marco regulatorio nacional se complementa con la política "Hoja de Ruta para el Desarrollo y Aplicación de la Inteligencia Artificial en Colombia"[1], formulada por el Ministerio de Ciencia, Tecnología e Innovación, que resalta la necesidad de aplicar principios éticos en el diseño de algoritmos, garantizar la transparencia y reducir los sesgos en los sistemas automatizados. En línea con estas directrices, el Banco Interamericano de Desarrollo[2] ha recomendado incorporar herramientas de IA en todas las etapas de diseño, ejecución y evaluación de políticas públicas, así como capacitar al talento humano en el uso responsable y estratégico de dichas herramientas.

En Colombia, aunque la política pública ha avanzado en la definición de lineamientos para la IA, los casos concretos de innovación pública con aplicación en el sector salud siguen siendo limitados. Este contexto plantea una oportunidad estratégica para que la ADRES

[1] Ministerio de Ciencia, Tecnología e Innovación (2024) Hoja de Ruta para el Desarrollo y Aplicación de la Inteligencia Artificial en Colombia. República de Colombia.

[2] Banco Interamericano de Desarrollo (2025) Artificial Intelligence Framework for the Inter-American Development Bank Group. Inter-American Development Bank. Disponible aquí.

lidere la incorporación de inteligencia artificial en el proceso de auditoría médica, fortaleciendo la capacidad institucional, mejorando la gestión del conocimiento y contribuyendo al cumplimiento de los objetivos de eficiencia, transparencia y sostenibilidad del sistema.

La implementación de soluciones basadas en IA, además de responder a las políticas nacionales de transformación digital, contribuirá a posicionar a la ADRES como una entidad líder en innovación pública dentro del sector salud. Al aprovechar sus capacidades técnicas y su experiencia institucional, la entidad podrá avanzar hacia una auditoría más ágil, precisa y fundamentada en evidencia, consolidando una administración moderna de los recursos públicos de la salud y fortaleciendo la confianza ciudadana en la gestión estatal.

3. Aspectos generales de la prueba de concepto.

Así las cosas, teniendo en cuenta la complejidad del reto de innovación pública que se plantea en el presente documento, y las oportunidades de desarrollo e implementación de nuevas soluciones tecnológicas que ofrece la inteligencia artificial, las cuales han sido identificadas y promovidas por los distintos documentos de política pública nacional en materia de transformación digital del Estado, la ADRES decidió formular una Prueba de Concepto (PoC, por sus siglas en inglés) que buscaba analizar la viabilidad de desarrollar e implementar un sistema de inteligencia artificial para apoyar el proceso de auditorías médicas para reclamaciones asociadas a accidentes de tránsito no SOAT. Esta categoría de reclamaciones se eligió como muestra inicial para la Prueba de Concepto (PoC), teniendo como objetivo: Evaluar la viabilidad de su implementación en una parte del proceso de auditoría y, al mismo tiempo, analizar su escalabilidad para una posible contratación futura.

- **Objetivo Estratégico:** El objetivo general de la PoC es "analizar la viabilidad técnica, normativa y económica de implementar un prototipo de inteligencia artificial (IA) que realice verificaciones automatizadas sobre los procesos de auditoría médica especializada en reclamaciones de cuentas asociadas a accidentes de tránsito".
- **Alcance de la PoC:** La PoC se ejecutó sin contraprestación económica para encontrar un Producto Mínimo Viable (PMV). El prototipo debía ser capaz de procesar información estructurada (RIPS, CUPS, CIE-10, etc.) y no estructurada (historias clínicas, normas, etc.) para la auditoría de reclamaciones por accidentes de tránsito no soat, aplicando las glosas definidas en el manual de la ADRES.

4. Proceso de convocatoria y vinculación de participantes

La invitación fue enviada el 09 de julio de 2025 y estuvo dirigida a las empresas Google Cloud, Amazon Web Services, Oracle y Microsoft Azure como los cuatro principales

prestadores de los servicios de tecnología descritos, cuyo criterio de selección, al no estar atado por ley a procesos de licitación pública, se basó en un estudio preliminar donde se encontró que estas empresas tenían una participación mayoritaria en el mercado de servicios tecnológicos con alojamiento en la nube y con la integración de soluciones basadas en inteligencia artificial. Estos cuatro prestadores respondieron a la invitación en alianza con las siguientes empresas ubicadas en el territorio colombiano:

- INFORMACIÓN LOCALIZADA S.A.S. _ Aliada de Google Cloud Platform
- HIGH TECH SOFTWARE S.A.S. _ Aliada de Amazon Web Services
- ENTELGY COLOMBIA S.A.S. _ Aliada de Oracle
- TI724 S.A.S. _ Aliada de Microsoft Azure.

Estas empresas enviaron los documentos solicitados en la invitación, así mismo, respondieron las comunicaciones a través de las cuales se solicitó la subsanación de los documentos para garantizar la transparencia, la legalidad y la igualdad de condiciones durante el ejercicio de la PoC. Estos documentos fueron:

- Hoja de vida institucional o perfil de la entidad
- Perfiles propuestos del equipo que participará en la PoC, donde debe incluirse formación y experiencia.
- Certificado de existencia o representación legal con un tiempo de expedición inferior a 30 días.
- Documento de identificación del Representante Legal
- Documentos que demuestren la existencia de una política institucional que prevenga, mitigue y contenga los riesgos implícitos en el manejo de datos personales e información sensible, el procesamiento de datos, y el almacenamiento seguro.
- Manifestación de interés debidamente diligenciada y firmada en el formato anexo a la presente invitación.
- Documento con descripción de la ubicación geográfica de los servidores o infraestructuras locales que alojan la información contenida en el servicio de nube de los participantes y sus aliados. Así como la normatividad aplicable a la gestión de datos asociada a dichos servidores o infraestructuras.
- Acuerdo de confidencialidad firmado por el representante legal del participante o del aliado estratégico que tenga domicilio en Colombia.

4.1. Manifestación de interés y firma de acuerdos de entendimiento.

Una vez manifestado el interés de participar en la PoC por parte de las empresas participantes, el 22 de julio de 2025 se firmaron tres documentos y se diligenció un inventario. Los documentos son los siguientes:

- Un Memorando de Entendimiento entre la ADRES y cada una de las empresas que manifestaron su interés de participar en la PoC.
- Un acuerdo de confidencialidad entre la ADRES y cada una de las empresas donde se fijan las condiciones del tratamiento de la información confidencial y aquella protegida por derechos de propiedad intelectual, la cual fue utilizada durante las actividades de la PoC.
- Un acuerdo de transferencia de datos entre la ADRES como responsable del tratamiento y cada una de las empresas como encargadas del tratamiento, donde se fijan las condiciones de tratamiento de los datos personales y los datos sensibles asociados a las cuentas médicas que conforman la muestra a partir de la cual se ejecutará la PoC.
- Un inventario de propiedad intelectual diligenciado por cada una de las empresas que participaron en la PoC, donde se identifican claramente los derechos de propiedad intelectual previos que pertenecen a la empresa y que serán usados durante la PoC.

El proceso de convocatoria y vinculación de participantes se desarrolló de conformidad con los criterios de selección objetiva establecidos en el artículo 5 de la Ley 1150 de 2007, específicamente aplicados al proceso de selección mediante la modalidad de contratación directa por la ejecución de actividades de ciencia, tecnología e innovación de acuerdo con el literal 3, numeral 4 artículo 2 de la Ley 1150 de 2007. Teniendo en cuenta especialmente los siguientes criterios:

- La capacidad jurídica, condiciones de experiencia, capacidad financiera y de organización de los proponentes.
- El uso de factores de calificación cuantitativos para valorar los aspectos técnicos y económicos del proyecto
- La definición de la oferta más favorable teniendo en cuenta los factores técnicos y económicos de escogencia a través de la ponderación precisa y detallada de los mismos.

Estas condiciones garantizaron el cumplimiento de los principios de transparencia, economía, responsabilidad e igualdad contractual, establecidos en la Ley 80 de 1993.

Así mismo, la suscripción de los memorandos de entendimiento, el acuerdo confidencialidad y el acuerdo de transferencia de datos personales, así como la elaboración del inventario de propiedad intelectual, permitió establecer un marco jurídico y operativo sólido para el desarrollo de la Prueba de Concepto (PoC).

De esta manera, la ADRES aseguró que todas las actuaciones estuvieran debidamente soportadas en instrumentos contractuales y administrativos que resguardaran la integridad de la información, la protección de los derechos de propiedad intelectual y el cumplimiento de las normas.

4.2. Cronograma de la prueba de concepto.

De acuerdo con las fases previstas durante la participación y el desarrollo de la Prueba de Concepto, se planteó el siguiente cronograma:

ID	TAREA	RESPONSABLE	DURACIÓN (días)	INICIO	FIN
1	FASE 1: Documentación			9-7-25	22-7-25
1.1	Envío invitación	ADRES	1	9-7-25	9-7-25
1.2	Envío manifestación de interés y documentos correspondiente:	PARTICIPANTES	7	9-7-25	15-7-25
1.3	Revisión/subsanación documentos	ADRES Y PARTICIPANTES	6	16-7-25	21-7-25
1.4	Evento firma del memorando de entendimiento	ADRES Y PARTICIPANTES	1	22-7-25	22-7-25
2	FASE 2: Primera Iteración			22-7-25	14-8-25
2.1	Entrega muestra de reclamaciones, fuentes estructuradas y no estructuradas	ADRES	1	22-7-25	22-7-25
2.2	Diseño y entrenamiento del prototipo IA	PARTICIPANTES	15	22-7-25	5-8-25
2.3	Apertura de canal de comunicación para dar respuesta a las inquietudes, socialización de resultados preliminares y retroalimentación.	ADRES Y PARTICIPANTES	22	23-7-25	13-8-25
2.4	Presentación y entrega 1 - informe + API consumible + Front	PARTICIPANTES	1	6-8-25	6-8-25
2.5	Testeo y análisis del prototipo IA	ADRES	8	6-8-25	13-8-25
2.6	Socialización de resultados preliminares y retroalimentación.	ADRES Y PARTICIPANTES	7	14-8-25	14-8-25
3	FASE 3: Segunda Iteración			7-8-25	22-8-25
3.1	Reentrenamiento y ajuste del prototipo IA	PARTICIPANTES	12	7-8-25	18-8-25
3.2	Entrega 2 - API consumible ajustada + Front	PARTICIPANTES	1	18-8-25	18-8-25
3.3	Testeo y análisis del prototipo IA	ADRES	4	19-8-25	22-8-25
3.4	Socialización de resultados preliminares y retroalimentación.	ADRES Y PARTICIPANTES	1	22-8-25	22-8-25
4	FASE 4: Tercera Iteración			19-8-25	3-9-25
4.1	Reentrenamiento y ajuste del prototipo IA	PARTICIPANTES	8	19-8-25	26-8-25
4.2	Entrega 3 - API consumible ajustada + Front	PARTICIPANTES	1	28-8-25	28-8-25
4.3	Testeo y análisis del prototipo IA	ADRES	6	28-8-25	2-9-25
4.4	Socialización de resultados preliminares y retroalimentación.	ADRES Y PARTICIPANTES	1	3-9-25	3-9-25
5	FASE 5: Cuarta Iteración			28-8-25	23-9-25
5.1	Reentrenamiento y ajuste del prototipo IA	PARTICIPANTES	19	28-8-25	15-9-25
5.2	Mesa de trabajo con equipo auditor	ADRES Y PARTICIPANTES	1	5-9-25	5-9-25
5.3	**Entrega de informe**	**PARTICIPANTES**	1	9-9-25	9-9-25
5.4	Espacios de trabajo con el equipo auditor	ADRES Y PARTICIPANTES	8	5-9-25	12-9-25
5.5	Entrega 4 - API consumible ajustada + Front	PARTICIPANTES	1	15-9-25	15-9-25
5.6	Testeo y análisis del prototipo IA	ADRES	7	15-9-25	21-9-25
5.7	Socialización de resultados preliminares y retroalimentación.	ADRES Y PARTICIPANTES	1	23-9-25	23-9-25
6	FASE 6: Entrega Producto Mínimo Viable			16-9-25	16-10-25
6.1	Reentrenamiento y ajuste final del prototipo IA	PARTICIPANTES	14	16-9-25	29-9-25
6.2	Espacios de trabajo con el equipo auditor	ADRES Y PARTICIPANTES	3	24-9-25	26-9-25
6.3	Presentación final PMV (Presencial)	ADRES Y PARTICIPANTES	3	30-9-25	2-10-25
6.4	Testeo y análisis final del prototipo IA	ADRES	8	3-10-25	10-10-25
6.5	Socialización de resultados finales PoC (SIA)	ADRES	1	14-10-25	14-10-25
	Socialización de resultados finales PoC (DOP-SIA)	ADRES	1	14-10-25	14-10-25
	Socialización de resultados finales PoC (Directores)	ADRES	1	15-10-25	15-10-25
6,6	Resultados nubes	ADRES Y PARTICIPANTES	1	16-10-25	16-10-25
6,7	Cierre de la PoC	ADRES Y PARTICIPANTES	1	20-10-25	20-10-25

5. Metodología de la prueba de concepto y estrategias de acompañamiento.

5.1. Planteamiento metodológico en desarrollo de la PoC.

El desarrollo de la PoC implicó el desarrollo de un prototipo que pudiera procesar datos estructurados (RNEC, RIPS, CUPS, CIE-10, REPS, RETHUS, manuales tarifarios) y no estructurados (normas legales, protocolos clínicos, historias clínicas digitalizadas, Manual Único de Glosas) en un sistema de inteligencia artificial que pueda emitir recomendaciones trazables y sustentadas en normativa vigente, y generando alertas automáticas ante posibles inconsistencias o glosas en cada una de las cuentas médicas.

El prototipo propuesto por las empresas dentro de la PoC debía cumplir requerimientos funcionales (procesamiento masivo, interpretación normativa automática, clasificación de glosas, integración con bases externas, interfaz de visualización, soporte a múltiples formatos, escalabilidad) y no funcionales (ciberseguridad con estándares ISO 27001 y NIST, protección de datos sensibles conforme a Ley 1581 de 2012, alta disponibilidad, principios éticos, control tecnológico por parte de ADRES, documentación y transferencia de conocimiento).

El desarrollo inicial de esta PoC se realizó en **fases iterativas** (cuatro ciclos de entrenamiento, ajuste y prueba), hasta obtener una quinta iteración donde se evalúa de manera definitiva el **Producto Mínimo Viable (PMV)** presentado por cada participante.

Los PMV presentados por cada una de las participantes fueron probados y evaluados por el equipo técnico de la ADRES en un entorno de pruebas controlado dispuesto por cada una de las participantes. Dicha evaluación buscaba analizar la viabilidad técnica, normativa y económica de implementar la solución tecnológica en entornos reales asociados al proceso de Auditoría de la ADRES, así como las capacidades y la idoneidad de las empresas participantes en el diseño del prototipo, con miras a explorar el posible codesarrollo, escalamiento e implementación de la solución tecnológica que obtenga la mejor calificación durante la PoC, a través de un Convenio Especial de Cooperación como posible mecanismo de asociación contractual.

Con este enfoque de desarrollo conjunto y transferencia de resultados, la PoC buscó fortalecer las capacidades institucionales de la ADRES, promover la soberanía tecnológica y garantizar el cumplimiento normativo, ético y técnico, impulsando una solución sostenible para la gestión pública en salud sin generar erogaciones presupuestales adicionales. La propuesta de Prueba de Concepto (PoC) buscó, por tanto, generar una arquitectura híbrida basada en inteligencia artificial entrenada con datos reales del sistema colombiano para reflejar con mayor fidelidad sus particularidades históricas, semánticas y regulatorias.

5.2. Entregables esperados.

Los **entregables** incluyeron:

- Un prototipo funcional de IA, API/REST que acoto las características técnicas planteadas en la PoC, y por la cual se realizaron los cargues masivos de reclamaciones durante las fases de testeo de la PoC.
- Un visor web de reclamaciones o FRONT que permitió la interacción del perfil de auditor con el análisis de las reclamaciones realizado por la IA .
- Informes parciales y final de ejecución con la metodología usada para la construcción de la solución tecnológica.

- Documentación técnica y de arquitectura.
- Plan de transferencia de resultados de la PoC con información sobre el código fuente creado, manuales y capacitación, y el inventario de resultados de la PoC que pueden llegar a ser protegibles por derechos de propiedad intelectual.

5.3. Metodología de evaluación.

La evaluación de los prototipos se basó en tres criterios ponderados, cuya calificación fue posteriormente ajustada por un factor de "cobertura".

La calificación base se compuso de la siguiente manera:

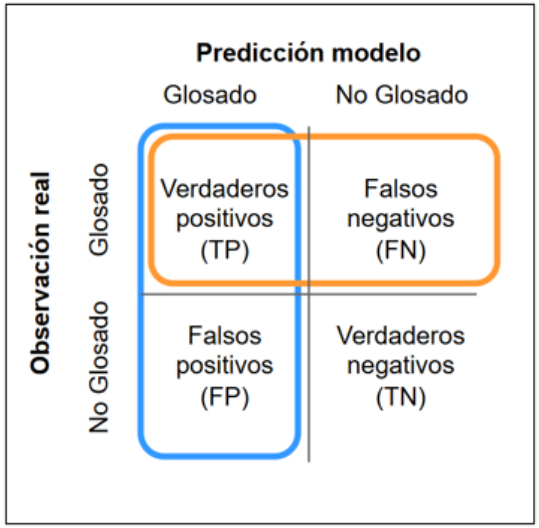
- **Precisión (50 %):** Evalúa que el prototipo pueda identificar, caracterizar y justificar glosas sobre cuentas médicas relacionadas con accidentes de tránsito, por falta de justificación de la pertinencia médica.
- **Eficiencia (35 %):** evalúa en función de los costos de inferencia proyectados para tres escenarios de producción (1,000, 10,000 y 100,000 reclamaciones en 24 horas), los cuales deberán ser calculados con un modelo de pago por uso (pay as you go) y soportados con los enlaces de las calculadoras públicas de cada escenario.
- **Desempeño (15 %):** Evalúa que el prototipo pueda generar recomendaciones técnicas que apoyen el juicio clínico y la toma de decisiones en los procesos de auditoría médica. Así mismo, se evalúa la coherencia en la justificación de la glosa, donde se verifica que las recomendaciones estén conectadas con los soportes técnicos y normativos debidamente referenciados.

  Este puntaje base fue ajustado por la **cobertura**, definida como el porcentaje de ítems que cada participante logró procesar sobre la muestra seleccionada para la evaluación del PMV.

Métricas de Evaluación Detalladas:

- **Medición de Precisión:** Evalúa que el prototipo pueda identificar, caracterizar y justificar glosas sobre cuentas médicas relacionadas con accidentes de tránsito, por falta de justificación de la pertinencia médica.

Matriz de confusión



TP: Glosas correctamente identificadas
FN: Glosas no identificadas por el modelo
FP: Reclamaciones que no eran glosas, pero el modelo las identificó como glosas
TN: Reclamaciones que no eran glosas, y el modelo las identificó correctamente

Métricas

$$Exactitud = \frac{TP}{TP + FP}$$

$$Sensibilidad = \frac{TP}{TP + FN}$$

$$Fbeta = \frac{(1 + \beta^2) * Exactitud * Sensibilidad}{(\beta^2 * Exactitud) + Sensibilidad}$$

La métrica de la precisión se basa en la interacción de los VP, pero teniendo en cuenta los falsos positivos (FP) y los falsos negativos (FN)

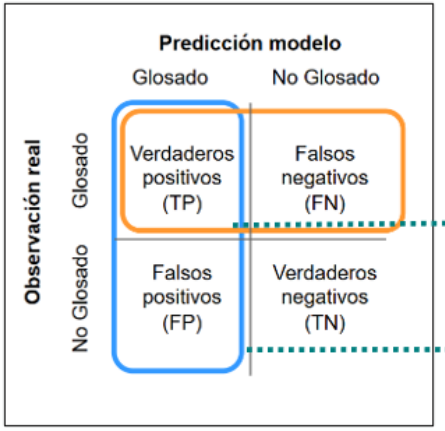
Consideraciones

- La unidad de análisis son los **ítems** de las reclamaciones glosados por justificación médica.
- La métrica se le da la mayor importancia los **VP,** lo que influye que el parámetro beta sea menor que 1 ($\beta < 1$). Para el PMV se fija $\boldsymbol{\beta = 0,635}$
- Se ajusta la métrica por la cobertura de procesamiento del PMV en las 24 horas

1 1 1 1

- **Medición de Desempeño:** Evalúa que el prototipo pueda generar recomendaciones técnicas que apoyen el juicio clínico y la toma de decisiones en los procesos de auditoría médica. Así mismo, se evalúa la coherencia en la justificación de la glosa, donde se verifica que las recomendaciones estén conectadas con los soportes técnicos y normativos debidamente referenciados.

Matriz de confusión



TP: Glosas correctamente identificadas
FN: Glosas no identificadas por el modelo
FP: Reclamaciones que no eran glosas, pero el modelo las identificó como glosas
TN: Reclamaciones que no eran glosas, y el modelo las identificó correctamente

Muestra ítems glosado nube

Las muestras fueron seleccionadas con un margen de error del 5% y 95% de confianza.

Muestra por nube de los **verdaderos positivos (TP),** se revisó la justificación de la glosa médica frente al resultado de la auditoria de la ADRES.

Muestra por nube de los **falsos positivos (FP),** se revisó los ítems no glosados por la ADRES, pero glosados por la nube debidamente soportados.

Consideraciones de revisión

- El equipo auditor revisó cada ítem glosado por nube, verificando la justificación de la glosa médica.
- Se estimo el criterio de desempeño extrapolando el resultado de la muestra.

16

El desempeño se calculó de la siguiente manera:

$$Desempeño = \frac{ítems\ justificados\ correctamente}{TP + FP}$$

Medición de Eficiencia: evalúa en función de los costos de inferencia proyectados para tres escenarios de producción (1,000, 10,000 y 100,000 reclamaciones en 24 horas), los cuales deberán ser calculados con un modelo de pago por uso (pay as you go) y soportados con los enlaces de las calculadoras públicas de cada escenario.

$$Eficiencia_{Total} = \frac{1}{3}\left(Eficiencia_{PoC} + Eficiencia_{Des_10mil} + Eficiencia_{Des_100mil}\right)$$

Interpretación:

- El menor costo reportado obtiene 100 puntos de eficiencia.
- El mayor costo obtiene una eficiencia menor proporcionalmente.
- Costos intermedios reciben una puntuación relativa, penalizada suavemente por el factor 0.9.

Donde:

- **Costo_[*nube*]**: es el costo pronosticado por el equipo evaluado para procesar las cuentas médicas en la nube.
- **Min (*Costo*)**: el costo más bajo entre todas las propuestas recibidas.
- **Max (*Costo*)**: el costo más alto entre todas las propuestas.
- **0.90**: factor de penalización para limitar el rango de variabilidad (evita que pequeñas diferencias se traduzcan en grandes penalizaciones).
- **1−…:** se invierte la escala, de modo que los **costos más bajos obtienen mayores eficiencias**.
- ***100**: convierte el resultado en porcentaje.

Para obtener el indicador de costos, se calculan los valores para tres escenarios (1.000, 10.000 y 100.000 reclamaciones en 24 horas) y, posteriormente, se promediaron los resultados.

$$Eficiencia_{Total} = \frac{1}{3}\left(Eficiencia_{PoC} + Eficiencia_{Des_10mil} + Eficiencia_{Des_100mil}\right)$$

Donde:

Calcula la **eficiencia total** como el **promedio aritmético** de tres eficiencias individuales:

1. $Eficiencia_{PoC}$:eficiencia de una etapa o proceso denominado *PoC*
2. $Eficiencia_{Des_10ml}$:eficiencia de un proceso de Despliegue en producción con 10.000 reclamaciones.
3. $Eficiencia_{Des_100ml}$:la misma medición anterior pero con 100.000 reclamaciones.

El factor $\frac{1}{3}$significa que las tres eficiencias tienen **el mismo peso** en el cálculo.

5.4. Estrategia de acompañamiento por parte del equipo auditor de ADRES

Para facilitar el entendimiento y el entrenamiento de los participantes, la ADRES estableció sesiones de retroalimentación después de cada iteración. En dichos encuentros, el equipo de auditoría médica de la ADRES presentaba los hallazgos de cada entrega y revisaba casos específicos, proveyendo observaciones para que cada "nube" mejorara su proceso.

En el marco de la PoC, y tras las Iteraciones 1 y 2, se hizo evidente que el desempeño inicial se enfocaba en el ámbito administrativo en lugar de la justificación médica. A raíz de esto, la ADRES implementó una metodología de acompañamiento específico.

Esta metodología comenzó con una capacitación en conceptos técnicos de auditoría médica ofrecida por la DOP (Dirección de Otras Prestaciones). El objetivo era clarificar los criterios que dan lugar a una glosa u objeción y la forma correcta de sustentar estos puntos en los soportes documentales.

Posteriormente, se establecieron espacios de acompañamiento personalizado con cada "nube" para entrenar a su Inteligencia Artificial (IA) en la generación de respuestas pertinentes al ámbito de la auditoría. Dicho entrenamiento se centró en modelos de aprendizaje automático para identificar patrones y algoritmos para comparar diagnósticos, tratamientos y procedimientos.

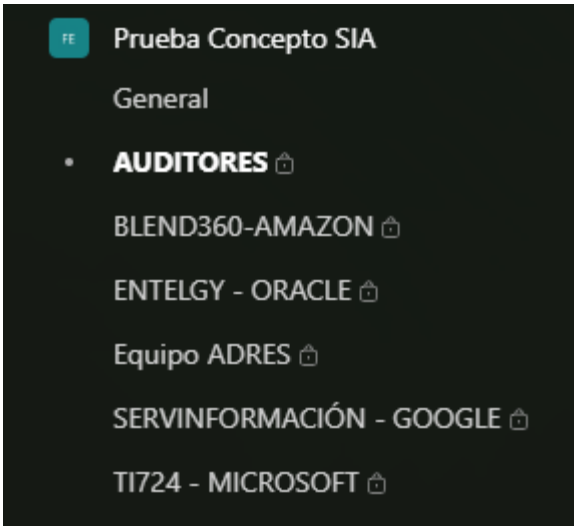
Se hizo especial énfasis en el cruce de datos de los soportes radicados por cada prestador y se concluyó resaltando la importancia de generar alertas y sugerencias para el cierre del proceso de auditoría médica.

Entre los insumos compartidos con los participantes se incluyeron metodologías para:

- La recopilación y organización de documentos.
- El análisis preliminar de la cuenta.
- La verificación detallada de facturación y registros.
- La elaboración del informe de hallazgos y ajustes.

5.5. Estrategia de comunicaciones y notificaciones

Como canal oficial para la Prueba de Concepto (PoC), se implementó un espacio de trabajo en Microsoft Teams. La finalidad fue doble: primero, asegurar una trazabilidad rigurosa de toda la comunicación y, segundo, mantener la confidencialidad de la información intercambiada con cada proveedor.



Para lograrlo, se configuró un equipo principal ("Prueba Concepto SIA") que contenía canales privados y de acceso restringido para cada aliado, tal como lo ilustran los nombres de canal con el ícono de candado (🔒) en la imagen (p.ej., "SERVINFORMACIÓN - GOOGLE", "TI724 - MICROSOFT"). Esta metodología permitió una gestión de proyecto ordenada y segura, facilitando el diálogo privado con cada participante.

Comunicaciones enviadas por Canal:

Aliado	*Blend 360 - Amazon*	*Entelgy - Oracle*	*Servonformación - Google*	*TI724 - Microsoft*
Comunicaciones Adres	53	54	56	51
Comunicaciones Aliado	35	19	11	12

*Cada comunicación con las réplicas correspondientes por el aliado y por el equipo de la adres

6. Concepto de seguridad de la información emitido por el oficial de riesgo de ADRES

Se realizó un análisis comparativo de las plataformas Hyperscaler participantes en la PoC, consultando su información pública sobre los siguientes aspectos de seguridad de la información y ciberseguridad:

- Gestión, detección y visualización de amenazas.
- Gestión de activos de ciberseguridad y riesgos asociados.
- Postura de seguridad de la plataforma.
- Cumplimiento normativo.
- Clasificación, economización y encapsulación de datos.
- Minimización del riesgo de exfiltración de datos.

Categoría	*Oracle OCI*	*Google GCP*	*Amazon AWS*	*Microsoft Azure*
Detección y visibilización de amenazas	Cloud Guard, Threat Intelligence Service, monitoreo continuo, alertas automatizadas.	Security Command Center, Event Threat Detection, Cloud Armor, AI Protection.	GuardDuty, Security Hub, AI/ML detección, correlación de alertas.	Defender for Cloud, Sentinel, Entra ID Protection, Microsoft Defender XDR, correlación de señales multicloud.
Gestión de activos y riesgos asociados	Compartimentos, rSTAR, aislamiento de recursos, cifrado por defecto.	Cloud SCC, DSPM, IAM, auditoría continua, clasificación automática.	AWS Config, Control Tower, IAM, Trusted Advisor, auditoría continua.	Defender for Cloud Inventory, Azure Resource Graph, Security Exposure Management, clasificación crítica automatizada.
Postura de seguridad	Seguridad por diseño, cifrado en tránsito/reposo, aislamiento físico/lógico.	Postura definida por políticas, monitoreo continuo, cumplimiento automatizado.	CSPM con Security Hub, cumplimiento con NIST, PCI, CIS.	CSPM multicloud, Zero Trust, cumplimiento automatizado, políticas como código, Secure Score.
Cumplimiento normativo	ISO 27001, SOC 1/2/3, PCI DSS, FedRAMP, C5.	GDPR, HIPAA, PCI DSS, ISO 27001, NIST 800-53.	SOC 2, ISO 27001, FedRAMP, HIPAA, GDPR, FISMA.	ISO 27001, NIST, PCI DSS, FedRAMP, GDPR, HIPAA, Microsoft Cloud Security Benchmark.
Clasificación y economización de datos	Etiquetas, políticas de ciclo de vida, análisis de uso, automatización de costos.	Cloud DLP, etiquetas, FinOps Hub, almacenamiento por clases.	Cost Explorer, Budgets, Compute Optimizer, clasificación por uso.	Azure Advisor, Cost Management, Turbonomic, etiquetado, políticas de

				retención, tiering automático.
Encapsulación de datos	Cifrado AES-256, TLS 1.3, Oracle Vault, separación lógica/física.	Confidential Computing, CMEK, TLS, IAM granular.	KMS, TLS, S3 encryption, IAM, VPC endpoints.	Cifrado en tránsito/reposo, Azure Key Vault, Private Link, VNet, NSG, ExpressRoute.
Minimización del riesgo de exfiltración	Monitoreo de actividad, rotación de claves, aislamiento de red.	VPC Service Controls, DLP API, IAM Conditions, auditoría.	DLP, IAM restrictivo, CloudTrail, reglas de exfiltración QRadar.	CAAC, Private Endpoints, políticas de servicio, monitoreo de tráfico, segmentación de red.

Integración con Centro de Operaciones de Seguridad (SOC)

Adicionalmente, se validó la capacidad de integración de las plataformas con el SOC del cliente. Todas las plataformas permiten la integración con un SOC externo y ofrecen herramientas nativas robustas para operaciones de seguridad.

Conclusión del Área de Seguridad de la Información:

"Una vez consultado y revisado estos aspectos se encuentra que todas las plataformas participantes tienen suficiente infraestructura, herramientas y estrategias asociadas a la seguridad de la información para el desarrollo del ejercicio de la POC - IA".

7. Desarrollo de la Prueba de Concepto

7.1. Metodología Técnica y Desempeño de Infraestructura

Para garantizar la evaluación objetiva y estandarizada de las soluciones, el equipo técnico de ADRES lideró un proceso riguroso de cargue y consolidación de resultados.

Inicialmente, se definió un **JSON Schema** (versiones 1.0 y 2.0) en conjunto con todos los participantes. Este esquema fue fundamental para estandarizar la estructura de respuesta de las APIs, asegurando que todos los proveedores entregaran el análisis de glosas tanto a nivel de reclamación como a nivel de ítem, incluyendo justificaciones y referencias documentales.

El proceso de evaluación en cada iteración siguió tres pasos técnicos:

1. **Preparación y Empaquetado:** Se generaron paquetes ZIP individuales por reclamación, conteniendo los soportes PDF, la factura XML y el detalle en Excel.
2. **Envío y Monitoreo:** Se cargaron los datos a las APIs de cada proveedor, iniciando un periodo de 24 horas para el procesamiento.
3. **Consulta y Estructuración:** Se consultaron los resultados y se consolidaron en tablas unificadas para el análisis de auditoría.

Un hallazgo técnico clave durante las primeras fases fue la madurez de la arquitectura de las API. Inicialmente, algunos participantes (como Google en la iteración 1 y Oracle) proveyeron APIs **síncronas**, que requerían esperar el procesamiento completo de una reclamación antes de enviar la siguiente, siendo un modelo lento. A partir de la iteración 3, AWS, Google y Microsoft migraron a APIs **asíncronas** basadas en colas, un modelo mucho más robusto y escalable.

Este enfoque técnico expuso desafíos significativos de infraestructura en varios participantes:

- **Google** presentó un límite técnico persistente, rechazando el cargue de archivos con un peso superior a 32.5 MB.
- **Oracle** experimentó fallos continuos en sus APIs síncronas durante las Fases 2 y 3, reportando errores de *timeout* y fallos en servicios externos (OCR), lo que impidió realizar las pruebas de cargue en esas iteraciones.
- **Microsoft** sufrió una caída en su sistema de colas durante la Fase 4, lo que resultó en una entrega de resultados muy baja (227 de 1363 reclamaciones).

En la fase final del **PMV (Sección 7.4)**, esta medición de infraestructura fue crítica para entender la capacidad real de procesamiento ("Cobertura") de cada proveedor sobre la muestra de 1.106 reclamaciones:

- **AWS:** Recibió 1.086 reclamaciones y procesó **el 100% de los ítems (51.318)** de dichas reclamaciones.
- **Google:** Recibió 1.036 reclamaciones, pero solo procesó el **86% de los ítems** de las mismas.
- **Microsoft:** Recibió 540 reclamaciones y solo procesó el **75% de los ítems** de las mismas.
- **Oracle:** Recibió 357 reclamaciones y solo procesó el **13% de los ítems** de las mismas, demostrando la menor capacidad de procesamiento.

7.2. Historial de Participación en las Iteraciones

	Iteración 1	*Iteración 2*	*Iteración 3*	*Iteración 4*	*Iteración 5(PMV)*
Servinformación - Google	Entrego un API y Front funcionales	Entrego un API y Front funcionales	Entrego un API y Front funcionales	Entrego un API y Front funcionales	Entrego un API y Front funcionales
Blend 360 – AWS	Entrego un API y Front funcionales	Entrego un API y Front funcionales	Entrego un API y Front funcionales	Entrego un API y Front funcionales	Entrego un API y Front funcionales
Entelgy – Oracle	Presento un primer acercamiento al entendimiento del negocio, pero sin API y Front funcionales	No realiza entrega	Realiza entrega, pero su solución API no fue funcional y no dio respuesta a la notificación del fallo.	Entrego un API y Front funcionales	Entrego un API y Front funcionales
TI 724 - Microsoft	Solo se presentó al espacio para solicitar tiempo adicional para realizar la primera entrega.	No realiza entrega	No realiza entrega	Entrego un API y Front funcionales	Entrego un API y Front funcionales

7.3. Resultados preliminares durante las primeras cuatro iteraciones.

A continuación, se detalla la evolución y los resultados finales de cada proveedor.

- **7.3.1. Proveedor: Google (Aliado: SERVINFORMACIÓN)**
    - **Iteración 1:** Procesó 446 de 1000 reclamaciones. A nivel de reclamación, mostró alta sensibilidad (99%) pero muy baja especificidad (1%), con una precisión del 80%. A nivel de ítem, el desempeño fue muy bajo (F-beta de 0.020). Críticamente, el **92.4% de las reclamaciones tuvo 0% de coincidencia** con los códigos de glosa correctos.
    - **Iteración 2:** El modelo mostró una sensibilidad del 100% pero una precisión del 73%, manteniendo la incapacidad de identificar casos sin glosa (especificidad de 0). La coincidencia de códigos de glosa seguía siendo un problema grave, con un 98.6% de los casos sin ninguna coincidencia.
    - **Iteración 3:** Mostró una mejora significativa. El F-beta a nivel de ítem subió a **0.085**. La coincidencia de códigos mejoró notablemente: los casos con 0% de

coincidencia bajaron al 66,4%, y un **18% de las reclamaciones superó el 60% de coincidencia**.

- **Iteración 4:**
  - **Rendimiento API:** Procesó 1.253 de 1.363 reclamaciones (91,9%).
  - **Precisión (Reclamación):** Excelente balance con **sensibilidad del 90,0%**, **precisión del 90,8%** y **F1-Score de 90,4%**.
  - **Precisión (Ítem):** Precisión perfecta (100%) pero sensibilidad baja (18,8%).
  - **Coincidencia de Códigos:** Sólido desempeño, con un **52,2% de las reclamaciones superando el 50% de coincidencia**.

En la justificación de glosas médicas, **Google inicialmente presentó un análisis enfocado en el ámbito administrativo**, lo cual limitó la evaluación correcta de dichas justificaciones. No obstante, a partir de la **tercera iteración**, se observó una **mejora progresiva en los resultados**. A pesar de esta mejora, la Prueba de Concepto finalizó con un número significativo de **referencias incorrectas**.

- **7.3.2. Proveedor: AWS (Aliado: Blend 360)**
  - **Iteración 1:** Procesó 976 de 1000 reclamaciones. El desempeño a nivel de reclamación fue similar al de Google, con alta sensibilidad (98%) pero muy baja especificidad (4%). El F-beta a nivel de ítem fue extremadamente bajo (0.011). La coincidencia de códigos fue casi inexistente: **95% de las reclamaciones tuvo 0% de coincidencia**.
  - **Iteración 2:** Mantuvo una sensibilidad del 100% pero con una especificidad de 0. La coincidencia de códigos de glosa seguía siendo muy pobre, con un 87.4% de los casos sin ninguna coincidencia.
  - **Iteración 3:** La sensibilidad a nivel de reclamación bajó al 77%. El F-beta a nivel de ítem mejoró a **0.061**, pero seguía siendo bajo. La coincidencia de códigos no mejoró significativamente, con **86,3% de los casos en 0% de coincidencia**.
  - **Iteración 4:**
    - **Rendimiento API:** Procesó 1.213 de 1.363 reclamaciones (89,0%).
    - **Precisión (Reclamación): Precisión alta (89,7%)** pero **sensibilidad moderada (63,9%)**, con un F1-Score de 74,7%.

- **Precisión (Ítem):** Precisión perfecta (100%) pero **sensibilidad muy baja (11,5%)**.
- **Coincidencia de Códigos:** Dificultades sistemáticas, con un **60,8% de las reclamaciones con coincidencia nula**.

Durante el proceso de la Prueba de Concepto, **Amazon demostró una mejora significativa en la identificación de glosas de tipo médico**. Inicialmente, la plataforma solo identificaba glosas y alertas de carácter administrativo; sin embargo, evolucionó hasta ofrecer una **combinación de ambas con un enfoque primario en el ámbito médico y una mayor precisión**.

Las alertas generadas por la "nube" **reflejaron una aplicación mejorada de las recomendaciones** impartidas por el equipo, lo que se tradujo en una **optimización de los resultados esperados**.

- **7.3.3. Proveedor: Oracle (Aliado: Entelgy)**
  - **Evolución (Iteraciones 1-2):** No participó.
  - **Iteración 3:** Se unió a la PoC, pero la solución técnica (API) que presentó **no fue funcional**. La empresa fue notificada del fallo, pero no ofreció una solución.
  - **Iteración 4:**
    - **Rendimiento API:** Procesó 1.027 de 1.363 reclamaciones (75,3%).
    - **Precisión (Reclamación):** Buen desempeño con **sensibilidad de 74,5%**, **precisión de 90,3%** y un sólido **F1-Score de 81,6%**.
    - **Precisión (Ítem): No fue posible realizar el análisis** por incompatibilidades técnicas.
    - **Coincidencia de Códigos:** Desempeño bimodal. Un **38,4% de los casos superó el 50% de coincidencia**, pero un 46,9% tuvo coincidencia nula.

En el caso de **Oracle**, no fue posible realizar sesiones de retroalimentación por parte del equipo auditor médico durante la **primera, segunda y tercera iteración**, debido a la **ausencia de resultados** que justificaran dichos encuentros. Sin embargo, en la **cuarta iteración**, se obtuvieron los primeros resultados, los cuales permitieron **constatar la insuficiencia de precisión** en la identificación de las glosas de tipo médico solicitadas en la Prueba de Concepto.

- **7.3.4. Proveedor: Microsoft (Aliado: TI-724)**
    - **Evolución (Iteraciones 1-3):** No participó activamente.
    - **Iteración 4:**
        - **Rendimiento API:** Muy bajo, procesando solo **222 de 1.363 reclamaciones (16,3%)**.
        - **Precisión (Reclamación):** Desempeño deficiente con **sensibilidad muy baja (12,8%)** y un F1-Score de solo 22,0%.
        - **Precisión (Ítem): Sensibilidad críticamente baja (0,9%)**.
        - **Coincidencia de Códigos:** Desempeño casi nulo, con **98,5% de las reclamaciones con menos del 1% de coincidencia**.

El desempeño de **Microsoft** fue similar al de Oracle, ya que **no se obtuvieron resultados en las primeras tres iteraciones**, lo que impidió la retroalimentación por parte del equipo médico. Por consiguiente, solo se pudo analizar la **cuarta iteración**, cuyos resultados **no mostraron coincidencias en las glosas médicas**. Esto se debió a que la plataforma mantuvo un **enfoque exclusivamente en la revisión administrativa**, lo cual resultó inadecuado para los objetivos de la Prueba de Concepto.

7.4. Resultados de la Entrega final del Producto Mínimo Viable (PMV)

La evaluación del Producto Mínimo Viable (PMV) se centró en determinar la viabilidad y madurez de los modelos de inteligencia artificial propuestos por cuatro proveedores (AWS, Google, Microsoft y Oracle) para la auditoría médica de reclamaciones por accidentes de tránsito.

Muestra del PMV

El análisis se realizó sobre una muestra probabilística de reclamaciones presentadas ante la ADRES.

- **Universo y Filtros:** El universo inicial de 34.354 reclamaciones se filtró para incluir únicamente presentaciones nuevas de mayo de 2025, con valores superiores a 500.000 pesos y pertenecientes a las zonas 3 y 4. Esto redujo el universo objetivo a 7.407 reclamaciones, equivalentes a 247.816 ítems.
- **Muestra Final:** Se obtuvo una muestra representativa con un 95% de confianza y un 5% de margen de error, compuesta por 1.106 reclamaciones y 55.117 ítems.

- **Estratificación:** La muestra se estratificó por el número de ítems por reclamación, según la siguiente distribución:

Estrato	*Ítems mínimos*	*Ítems máximos*	*Universo (Reclamaciones)*	*Muestra (Reclamaciones)*
Estrato 1	1	4	1.587	17
Estrato 2	5	16	1.991	77
Estrato 3	17	66	2.655	406
Estrato 4	67	265	1.097	529
Estrato 5	268	1.08	77	77
Total	1	1.08	7.407	1.106

Resultados de Cobertura de Procesamiento

La cobertura midió la capacidad de cada proveedor para procesar y analizar los 55.117 ítems de la muestra.

- **AWS:** Alcanzó una lectura casi total de la muestra, procesando **53.296 ítems** (96,7% de cobertura). Se destacó que AWS revisó el 100% de los ítems en las reclamaciones que procesó.
- **Google:** Logró procesar **31.037 ítems** (56,3% de cobertura).
- **Microsoft:** Presentó una cobertura sustancialmente inferior, procesando **12.550 ítems** (22,8% de cobertura). Se observó que su arquitectura no fue lo suficientemente robusta o escalable para el procesamiento requerido.
- **Oracle:** Tuvo el desempeño más bajo, procesando solo **1.158 ítems** (2,1% de cobertura). Se reportó que Oracle entregó una API síncrona en lugar de una asíncrona, mostrando bajo rendimiento.

Resultados de Evaluación (Puntajes Normalizados)

Los resultados de cada proveedor se normalizaron a una escala de 0 a 100 para permitir una comparación homogénea, aplicando las ponderaciones definidas (Precisión 50%, Desempeño 15%, Eficiencia 35%).

Precisión (Ponderación: 50%)

El análisis de precisión se restringió a 74 códigos de glosas médicas. La matriz de confusión resultante a nivel de ítem fue la siguiente:

Proveedor	Cobertura	Accuracy	Recall	FBeta	Evaluación precisión normalizada
AWS	96,7	26,6	37,5	30,1	100,0
Google	56,3	15,5	39,2	12,5	41,4
Microsoft	22,8	11,5	0,2	0,1	0,0
Oracle	2,1	14,1	85,9	0,5	1,4

Al aplicar la métrica F-Beta (con beta = 0,635) y normalizar los resultados, AWS obtuvo el mejor desempeño. Google mostró una capacidad de detección regular, pero con una alta tasa de falsos positivos, mientras que Microsoft y Oracle presentaron desempeños marginales.

Desempeño (Ponderación: 15%)

Se evaluó la calidad de la justificación de las glosas mediante una revisión por parte del equipo auditor sobre muestras de Verdaderos Positivos (TP) y Falsos Positivos (FP).

- **Justificación de Verdaderos Positivos (TPs):**
  - AWS justificó correctamente el 59,9%.
  - Google justificó correctamente el 52,3%.
  - Oracle justificó correctamente el 11,3%.
  - Microsoft no presentó glosas correctamente justificadas.
- **Justificación de Falsos Positivos (FPs):**
  - Google logró justificar el 47,6% de los FP.
  - AWS logró justificar el 34,5% de los FP.

Proveedor	Verdaderos positivos			Falsos positivos			Glosado Nube			Desempeño normalizado
	Justificados	Glosados	Desempeño	Justificados	Glosados	Desempeño	Justificados	Glosados	Desempeño	
AWS	4.083	6.817	59,9	1.734	5.020	34,5	5.817	11.837	49,1	98,7
Google	2.744	5.249	52,3	2.912	6.111	47,6	5.655	11.360	49,8	100,0
Microsoft	-	1.088	-	-	2.127	-	-	3.215	-	-
Oracle	23	200	11,3	-	447	-	23	647	3,5	7,0

Eficiencia (Ponderación: 35%)

La eficiencia se midió según los costos proyectados en tres escenarios (1.000, 10.000 y 100.000 reclamaciones en 24 horas).

- **Resultados por Escenario (sin ajustar):** Oracle fue la opción más económica en los escenarios de 1.000 y 10.000 reclamaciones. Microsoft resultó ser el más eficiente en el escenario de 100.000 reclamaciones, reflejando economías de escala.
- **Resultados Normalizados (Ajustados por Cobertura):** Al aplicar el ajuste por la cobertura de procesamiento real, los resultados cambiaron significativamente. AWS se posicionó como la opción más eficiente en términos reales. No fue posible calcular una métrica confiable para Google debido a su limitada eficiencia.

Proveedor	Escenario 1 (1.000 diarias)	Escenario 2 (10.000 diarias)	Escenario 3 (100.000 diarias)	Eficiencia PoC	Eficiencia 10 mil	Eficiencia 100 mil	Eficiencia Total	Cobertura	Eficiencia Ajustada	Eficiencia normalizada
AWS	$9.718.605	$60.048.520	$574.673.718	25,0	27,1	27,1	26,4	96,7	25,5	100,0
Google	$12.238.245	$77.138.010	$751.842.585	2,8	1,0	1,0	1,6	56,3	0,9	-
Microsoft	$ 12.440.470	$ 20.229.030	$ 79.085.025	1,0	87,8	100,0	62,9	22,8	14,3	54,6
Oracle	$1.226.250	$12.262.200	$122.622.075	100,0	100,0	93,6	97,9	2,1	2,1	4,7

Evaluación Total y Conclusión del PMV

La calificación total se obtuvo mediante el promedio ponderado de los tres componentes.

Proveedor	Precisión (35%)		Desempeño (15%)		Eficiencia (35%)		Total
	Indicador	Normalizado	Indicador	Normalizado	Indicador	Normalizado	
AWS	30,1	100,0	49,1	98,7	25,5	100,0	99,8
Google	12,5	41,4	49,8	100,0	0,9	-	35,7
Microsoft	0,1	-	-	-	14,3	54,6	19,1
Oracle	0,5	1,4	3,5	7,0	2,1	4,7	3,4

Conclusión: AWS se posicionó como el proveedor con el mejor desempeño integral, alcanzando una calificación casi perfecta de **99,8 puntos**. Su rendimiento demostró un equilibrio entre alta precisión, excelente justificación de resultados y la mayor eficiencia económica ajustada a su capacidad operativa rea. La alta cobertura (96,7%) y la consistencia en todos los componentes evaluados confirman su idoneidad y madurez técnica para un escalamiento a producción.

Por su parte, Google mostró un desempeño medio, destacando por una adecuada capacidad de justificación, pero limitado por su cobertura efectiva y costos relativos. Microsoft solo presentó eficiencia en escenarios de alta demanda, y Oracle, aunque con bajos costos unitarios, no alcanzó los niveles de precisión ni cobertura suficientes.

8. Conclusiones y Recomendaciones

8.1. Conclusiones

1. **Viabilidad Confirmada:** La Prueba de Concepto (PoC) cumplió exitosamente su objetivo principal al "analizar la viabilidad técnica, normativa y económica" de un sistema de IA para la auditoría médica. Los resultados demuestran que la tecnología de arquitectura híbrida (LLM, RAG y reglas) puede procesar datos estructurados y no estructurados para identificar glosas de pertinencia médica de manera eficaz.

2. **El Acompañamiento fue Crítico:** El proceso iterativo demostró ser esencial. Los resultados iniciales de todos los proveedores fueron deficientes, con un enfoque erróneo en aspectos administrativos y una nula capacidad para identificar glosas médicas. La estrategia de acompañamiento, capacitación y retroalimentación liderada por el equipo auditor de ADRES fue el factor decisivo que permitió a los proveedores (especialmente a AWS y Google) reentrenar sus modelos y mejorar la precisión.

3. **Capacidad de un Proveedor:** La evaluación final del PMV identificó a **AWS (aliado Blend 360)** como el ganador de la PoC, con una calificación ponderada de **99,8/100**. Su solución destacó en las cuatro dimensiones:
    a. **Cobertura:** Procesó el 96,7% de la muestra.
    b. **Precisión:** Obtuvo el F-Beta más alto (30,1).
    c. **Desempeño:** Logró la mejor combinación de justificación de verdaderos y falsos positivos.
    d. **Eficiencia:** Presentó la mejor eficiencia ajustada por su cobertura real.

4. **Seguridad Garantizada:** El análisis de seguridad de la información determinó que las cuatro plataformas *Hyperscaler* (AWS, Google, Microsoft y Oracle) cuentan con la infraestructura, herramientas y estrategias suficientes para garantizar la seguridad, cumplimiento y protección de los datos sensibles de ADRES.

8.2. Recomendaciones

Con base en los hallazgos de la PoC, el equipo técnico y auditor de ADRES recomienda:

1. **Avanzar a la Fase de Escalamiento:** Iniciar gestiones para la siguiente fase del proyecto con el proveedor ganador, **Amazon Web Services (AWS) y su aliado Blend 360**, con miras a un "posible codesarrollo, escalamiento e implementación de la solución".

2. **Explorar el Mecanismo de Contratación:** Evaluar la viabilidad de utilizar un **Convenio Especial de Cooperación**, u otro mecanismo de contratación directa permitido para ciencia y tecnología, que permita el desarrollo conjunto y asegure la transferencia de conocimiento y soberanía tecnológica a la ADRES.
3. **Evaluar la Escalabilidad Futura:** Una vez consolidado el modelo en el dominio actual, realizar estudios para escalar la solución a otros procesos de auditoría de mayor complejidad y volumen dentro de la entidad.
4. **Fortalecer Capacidades Internas:** Continuar con la capacitación del talento humano de ADRES en el uso y la auditoría de estas nuevas herramientas de IA, asegurando que la entidad mantenga el control y la capacidad de validación sobre las decisiones automatizadas.