

Nombre del proyecto

Workshop: Impacto Social y Responsabilidad
--

Profesor	Materia	Fecha de Inicio	Fecha de Finalización
Ing. Gladys Villegas	Proyecto Integrador	12/10/2025	16/10/2025

Resumen del Proyecto

<p>El proyecto propone un sistema de inteligencia artificial capaz de analizar ecografías para detectar de forma temprana el cáncer de tiroides, reduciendo errores y desigualdades en el diagnóstico médico. Se identifican como actores principales a médicos, pacientes, instituciones de salud y desarrolladores, garantizando su participación ética y técnica.</p> <p>El sistema ofrece beneficios sociales al mejorar la precisión y accesibilidad del diagnóstico, aunque plantea riesgos como sesgos algorítmicos, pérdida de privacidad y dependencia tecnológica. Para mitigarlos, se establecen estrategias de auditoría, anonimización y <i>Explainable AI</i>. Finalmente, se promueve un marco ético y de responsabilidad compartida que prioriza la equidad, la transparencia y el bienestar del paciente.</p>
--

Responsables del proyecto

Nombre	Rol	Tarea
Christian Garcia	Estudiante	Workshop Impacto Social y Responsabilidad de tu proyecto
Byron Piedra	Estudiante	Workshop Impacto Social y Responsabilidad de tu proyecto

## **Parte 1: Contexto del Proyecto**

### **1.1 Descripción del Sistema**

#### **Problema que resuelve**

El cáncer de tiroides es una de las neoplasias endocrinas más comunes y su detección temprana depende en gran medida de la interpretación experta de ecografías, lo que genera limitaciones en regiones con escasez de especialistas. Este proyecto busca reducir los errores de diagnóstico, costos y los tiempos de espera mediante un sistema de inteligencia artificial capaz de analizar imágenes ecográficas y apoyar la toma de decisiones médicas.

#### **Funcionalidad principal**

El sistema utiliza modelos de aprendizaje profundo (Deep Learning) para procesar imágenes ecográficas de la glándula tiroides, identificar posibles nódulos sospechosos y clasificarlos según su probabilidad de malignidad. Los resultados se presentan en una interfaz clara que permite al médico visualizar las regiones de interés y obtener un informe automatizado de apoyo diagnóstico.

#### **Usuarios objetivo**

El sistema está dirigido principalmente a médicos radiólogos, endocrinólogos y profesionales de la salud involucrados en el diagnóstico de enfermedades tiroideas. También puede ser utilizado por centros médicos, hospitales regionales y laboratorios que busquen optimizar la detección temprana del cáncer de tiroides.

#### **Dominio de aplicación**

El proyecto se ubica en el campo de la salud digital, específicamente en el área de la inteligencia artificial aplicada a la imagenología médica (Computer-Aided Diagnosis – CAD) y la telemedicina.

### **1.2 Alcance de Implementación**

#### **¿Dónde se implementaría?**

Inicialmente se implementaría en hospitales y centros de diagnóstico con servicios de ecografía, priorizando instituciones públicas o clínicas universitarias que atienden a comunidades con limitado acceso a diagnósticos por imágenes médicas.

Posteriormente podría escalarse a nivel nacional mediante integración con sistemas de salud pública o plataformas de telemedicina.

#### **¿Cuántas personas afectaría?**

El sistema beneficiaría directamente a médicos y pacientes en etapa de diagnóstico, impactando potencialmente a cientos de personas por centro de salud al año. A

largo plazo, su adopción podría beneficiar a miles de pacientes, reduciendo diagnósticos tardíos, precios y mejorando las tasas de supervivencia.

### ¿Qué decisiones toma o informa?

El sistema no sustituye al médico, sino que informa y apoya decisiones clínicas, indicando la probabilidad de malignidad de un nódulo y sugiriendo la necesidad de estudios complementarios o biopsias. Facilita así una toma de decisiones más rápida, precisa y basada en evidencia.

## Parte 2: Análisis de Stakeholders

### Identificación de Grupos de Stakeholders

Stakeholder	Descripción	Beneficios	Riesgos	Poder/Voz
<b>Médicos radiólogos y endocrinólogos</b>	Profesionales responsables del diagnóstico y seguimiento de pacientes con nódulos tiroideos.	Mayor precisión diagnóstica, apoyo en la interpretación de imágenes y reducción del tiempo de análisis.	Dependencia excesiva del sistema o resistencia inicial al cambio tecnológico.	<b>Alto</b>
<b>Pacientes</b>	Personas que se someten a estudios ecográficos para diagnóstico o control de enfermedades tiroideas.	Diagnóstico más temprano y confiable, reducción de errores humanos y acceso a atención de calidad incluso en zonas remotas.	Riesgo de malinterpretación de resultados o temor por el uso de IA en decisiones médicas.	<b>Medio</b>
<b>Instituciones de salud (hospitales y clínicas)</b>	Centros que implementan el sistema para optimizar procesos y mejorar servicios.	Eficiencia operativa, prestigio institucional, reducción de costos por diagnósticos tardíos.	Costos iniciales de implementación y capacitación del personal.	<b>Alto</b>

Stakeholder	Descripción	Beneficios	Riesgos	Poder/Voz
<b>Equipo de desarrollo e investigadores</b>	Grupo encargado del diseño, entrenamiento y validación del modelo de IA.	Desarrollo de tecnología innovadora, publicaciones científicas y reconocimiento profesional.	Riesgos éticos y legales si el sistema produce errores o sesgos en el diagnóstico.	<b>Medio</b>

## Análisis Individual de Stakeholders

### 1. Médicos radiólogos y endocrinólogos

- **Impacto en su trabajo:** El sistema mejora la precisión diagnóstica y acelera el análisis de imágenes, facilitando decisiones clínicas más informadas.
- **Consideración en el diseño:** Se incorporaron criterios de diagnóstico TI-RADS y se realizaron consultas con especialistas para definir las características más relevantes.
- **Protección de intereses:** Se garantiza mediante entrenamiento continuo, control humano final y reportes explicativos que mantienen la autonomía del médico.

### 2. Pacientes

- **Impacto en su vida:** Permite diagnósticos más rápidos, reduciendo la ansiedad y el riesgo de tratamientos tardíos.
- **Consideración en el diseño:** Se priorizó la protección de datos personales, anonimización de imágenes y una interfaz que fomente la confianza médica.
- **Protección de intereses:** Se aplican normas éticas y de confidencialidad (HIPAA, RGPD) y se asegura que los resultados sean interpretados por un profesional.

### 3. Instituciones de salud (hospitales y clínicas)

- **Impacto en su operación:** Mejora la productividad de los servicios de diagnóstico y eleva la calidad asistencial.
- **Consideración en el diseño:** Se contemplaron requisitos técnicos de integración con sistemas hospitalarios (PACS y HIS) y facilidad de uso.

- **Protección de intereses:** Se ofrecen licencias seguras, soporte técnico y acuerdos de confidencialidad que resguardan la información institucional.

#### 4. Equipo de desarrollo e investigadores

- **Impacto en su trabajo:** Fomenta la innovación científica y el avance de la inteligencia artificial en la medicina.
- **Consideración en el diseño:** Se promueve la revisión por pares, la trazabilidad de los datos y la evaluación ética de los modelos.
- **Protección de intereses:** Se garantiza mediante propiedad intelectual reconocida, transparencia en resultados y cumplimiento de normativas éticas de investigación.

### Parte 3: Evaluación de Impacto Social

#### 3.1 Impactos Positivos

##### Mejora en la detección temprana del cáncer de tiroides

- **Descripción concreta:**  
El sistema de IA permite identificar de forma automática y temprana nódulos sospechosos en ecografías, reduciendo los casos de diagnóstico tardío.
- **Grupos beneficiados:**  
Pacientes con riesgo de cáncer de tiroides, médicos radiólogos y centros de salud.
- **Evidencia o justificación:**  
Estudios recientes muestran que los modelos de aprendizaje profundo alcanzan precisiones superiores al 90% en la clasificación de nódulos tiroideos, mejorando la capacidad de detección frente al diagnóstico visual tradicional.

##### Democratización del acceso a diagnósticos especializados

- **Descripción concreta:**  
El sistema puede implementarse en hospitales regionales o rurales, brindando apoyo diagnóstico sin necesidad de un especialista presencial.
- **Grupos beneficiados:**  
Poblaciones de zonas rurales y comunidades con limitada cobertura médica.
- **Evidencia o justificación:**  
La telemedicina apoyada por IA reduce la brecha en el acceso a atención médica especializada y disminuye las desigualdades en salud.

### Optimización de recursos en instituciones de salud

- **Descripción concreta:** Automatizar la preclasificación de imágenes disminuye la carga de trabajo de los radiólogos y optimiza la gestión del tiempo y los recursos hospitalarios.
- **Grupos beneficiados:** Instituciones públicas y privadas, personal médico y pacientes.
- **Evidencia o justificación:** En estudios hospitalarios, la implementación de sistemas CAD (Computer-Aided Diagnosis) ha reducido en un 30% el tiempo promedio de análisis por caso.

### 3.2 Impactos Negativos y Riesgos

#### [Riesgo/impacto negativo 1: Sesgo en el entrenamiento del modelo de IA]

- **Descripción del riesgo:** Si las imágenes de entrenamiento provienen mayoritariamente de un tipo de población (por ejemplo, un rango de edad o características fisiológicas específicas), el sistema podría presentar menor precisión en otros grupos.
- **Grupos vulnerables:** Mujeres jóvenes, adultos mayores y pacientes con enfermedades tiroideas atípicas.
- **Severidad:** Alta
- **Probabilidad:** Media

#### [Riesgo/impacto negativo 2: Dependencia tecnológica y pérdida de criterio clínico]

- **Descripción del riesgo:** El uso prolongado del sistema podría generar dependencia por parte de los médicos, reduciendo la práctica de análisis crítico frente a casos complejos.
- **Grupos vulnerables:** Médicos jóvenes o con menor experiencia.
- **Severidad:** Media
- **Probabilidad:** Alta

### [Riesgo/impacto negativo 3: Privacidad y seguridad de datos médicos]

- **Descripción del riesgo:** Las imágenes ecográficas y los datos personales de los pacientes podrían ser vulnerables ante ataques o filtraciones si no se aplican medidas de protección adecuadas.
- **Grupos vulnerables:** Pacientes y hospitales que almacenan información sensible.
- **Severidad:** Alta
- **Probabilidad:** Media

### 3.3 Análisis de Equidad

- **¿Tu sistema reduce o aumenta desigualdades existentes?**  
Reduce desigualdades al brindar acceso diagnóstico en comunidades con escasos recursos médicos, aunque su efectividad dependerá del acceso a equipos de ecografía y conexión a internet.
- **Distribución de beneficios:**  
Los mayores beneficiados son los pacientes y los médicos de zonas rurales. Las grandes instituciones también se benefician por la eficiencia operativa, aunque en menor medida.
- **Distribución de riesgos:**  
Los desarrolladores y las instituciones asumen los riesgos técnicos y legales asociados al uso del sistema; los pacientes podrían verse afectados por errores de predicción.
- **Acceso:**  
El sistema está diseñado para funcionar con equipos estándar de ultrasonido, facilitando su implementación en centros regionales. Sin embargo, el acceso digital desigual puede limitar su adopción.
- **Barreras:**  
Limitaciones tecnológicas, costos de implementación inicial y resistencia cultural o desconfianza hacia el uso de IA en la medicina podrían excluir a ciertos grupos.

### Parte 4: Riesgos Éticos Específicos

A continuación, se analizan cuatro riesgos éticos clave identificados en el proyecto, en relación con los principios de justicia, privacidad, transparencia y responsabilidad.

## 1. Riesgo Ético: Sesgo algorítmico (Fairness y Sesgo)

**Descripción:** El modelo de IA podría discriminar o presentar menor precisión en determinados grupos poblacionales si los datos de entrenamiento no incluyen suficiente diversidad (por ejemplo, variaciones de edad, género o características fisiológicas de la tiroides).

**Evidencia:** Los estudios muestran que modelos entrenados con datos clínicos de poblaciones limitadas tienden a generalizar mal cuando se aplican a nuevas regiones o etnias, comprometiendo la equidad diagnóstica.

**Severidad:**Alta

**Grupo afectado:** Pacientes de poblaciones minoritarias o con características anatómicas menos representadas.

**Medidas de mitigación:**

- Balancear el dataset incluyendo imágenes de diversas fuentes geográficas y demográficas.
- Evaluar métricas de *fairness* y realizar auditorías de desempeño segmentadas.
- Mantener transparencia sobre las limitaciones del conjunto de datos.

## 2. Riesgo Ético: Privacidad y confidencialidad de datos médicos (Privacidad)

**Descripción:**

El sistema requiere almacenar y procesar imágenes ecográficas con información sensible, lo que plantea riesgos de filtración, reidentificación o uso no autorizado de datos personales.

**Evidencia:**

La manipulación de datos biomédicos sin protocolos de anonimización puede vulnerar regulaciones como la Ley Orgánica de Protección de Datos Personales (Ecuador) o el RGPD (UE).

**Severidad:** Alta

**Grupo afectado:** Pacientes y hospitales que comparten información médica.

**Medidas de mitigación:**

- Aplicar anonimización completa antes del entrenamiento del modelo.
- Uso de almacenamiento seguro y cifrado.
- Consentimiento informado explícito y trazabilidad de los datos.
- Auditorías periódicas del sistema.



### 3. Riesgo Ético: Falta de explicabilidad del modelo (Transparencia)

**Descripción:**

El sistema de IA puede generar predicciones sin una explicación comprensible para los médicos o pacientes, lo que limita la confianza y la adopción del sistema.

**Evidencia:**

Los modelos de *deep learning* son frecuentemente considerados “cajas negras”, dificultando la interpretación clínica de los resultados.

**Severidad:** Media

**Grupo afectado:** Médicos y pacientes que reciben información diagnóstica sin contexto.

**Medidas de mitigación:**

- Implementar mecanismos de *explainable AI* (por ejemplo, mapas de calor sobre regiones de interés).
- Incluir documentación accesible y mensajes claros sobre las limitaciones del modelo.
- Proporcionar capacitación al personal médico sobre la interpretación de los resultados.

### 4. Riesgo Ético: Responsabilidad ante errores diagnósticos (Accountability y Seguridad)

**Descripción:**

Si el modelo falla en la predicción (falso positivo o negativo), es necesario definir claramente quién asume la responsabilidad y cómo se gestionan los errores.

**Evidencia:**

En entornos clínicos, un error de clasificación podría derivar en tratamientos innecesarios o en la omisión de un diagnóstico temprano, afectando directamente la salud del paciente.

**Severidad:** Alta

**Grupo afectado:** Médicos, pacientes y la institución de salud que implementa el sistema.

## Parte 5: Estrategias de Mitigación

### Riesgo a mitigar: Sesgo algorítmico en el modelo

#### Estrategia propuesta:

Implementar un proceso continuo de evaluación de métricas de fairness y balanceo de datos durante el entrenamiento, asegurando que el modelo mantenga precisión equivalente entre distintos grupos poblacionales.

**Tipo:** Técnica

#### Implementación:

1. Aplicar métricas de equidad (Demographic Parity y Equal Opportunity) en cada ciclo de entrenamiento.
2. Ampliar el conjunto de datos incluyendo imágenes ecográficas de distintas regiones, edades y géneros.
3. Ajustar umbrales de decisión mediante *post-processing* cuando se detecten desviaciones significativas.

**Cuándo:** Antes del lanzamiento y de forma trimestral durante mantenimiento.

**Quién es responsable:** Equipo de *data science* y líder técnico del proyecto.

**Efectividad esperada:** Alta

### Riesgo a mitigar: Vulneración de privacidad y confidencialidad de datos médicos

#### Estrategia propuesta:

Aplicar anonimización, cifrado y control de acceso basado en roles para todos los datos utilizados por el sistema, conforme a las normas de protección de datos médicas y éticas.

**Tipo:** Política / Técnica

#### Implementación:

1. Implementar algoritmos de anonimización y eliminación de metadatos personales antes del almacenamiento.
2. Establecer políticas de acceso seguro y registro de auditorías sobre uso de datos.
3. Realizar revisiones de seguridad y auditorías internas cada seis meses.

**Cuándo:** Antes del despliegue y de forma continua durante la operación.

**Quién es responsable:** Administrador de seguridad y comité ético del proyecto.

**Efectividad esperada:** Alta

### **Riesgo a mitigar: Falta de explicabilidad del modelo**

#### **Estrategia propuesta:**

Integrar técnicas de *Explainable AI (XAI)*, como mapas de activación visual (*Grad-CAM*) y reportes interpretativos, para que los médicos comprendan las razones detrás de cada predicción.

**Tipo:** Diseño / Educación

#### **Implementación:**

1. Incorporar visualizaciones que resalten las áreas del ultrasonido utilizadas en la predicción.
2. Incluir en la interfaz una sección de “confianza del modelo” con niveles de probabilidad y advertencias.
3. Capacitar a los médicos en el uso e interpretación de las salidas del sistema.

**Cuándo:** Durante el desarrollo del prototipo y antes del lanzamiento oficial.

**Quién es responsable:** Equipo de diseño UX y líder clínico del proyecto.

**Efectividad esperada:** Media-Alta

### **Riesgo a mitigar: Falta de responsabilidad y trazabilidad ante errores diagnósticos**

#### **Estrategia propuesta:**

Definir un marco de responsabilidad compartida entre el sistema, los desarrolladores y el personal médico, junto con un registro de decisiones y trazabilidad de cada diagnóstico generado.

**Tipo:** Política / Gobernanza

#### **Implementación:**

1. Incorporar trazabilidad completa de cada predicción (imagen, modelo, versión y resultado).
2. Establecer un protocolo de revisión clínica obligatoria antes de emitir diagnósticos finales.
3. Crear un comité de auditoría ética que evalúe errores o apelaciones de casos dudosos.

**Cuándo:** Antes de la puesta en marcha y de manera continua durante la operación.

**Quién es responsable:** Comité de ética médica y dirección del proyecto.

**Efectividad esperada:** Alta

## Parte 6: Framework de Responsabilidad

### 6.1 Cadena de Responsabilidad

Rol	Responsabilidades	Rendición de Cuentas
<b>Desarrolladores</b>	Implementar el modelo de IA garantizando precisión, seguridad del código y control de versiones.	Pruebas unitarias, validación cruzada, revisión de código ( <i>code reviews</i> ).
<b>Data Scientists</b>	Asegurar la calidad, balance y equidad de los datos; evaluar métricas de <i>fairness</i> y sesgo.	Auditorías trimestrales de sesgo, documentación de métricas y resultados.
<b>Product Owner / Líder del proyecto</b>	Supervisar decisiones de diseño, priorización de funcionalidades y cumplimiento ético.	Documentación de decisiones, gestión de riesgos y seguimiento de KPIs éticos.
<b>Organización / Institución de salud</b>	Definir políticas de gobernanza, protección de datos y supervisión clínica.	Políticas de privacidad, reportes de auditoría y control de cumplimiento normativo.
<b>Comité Ético y Médico Asesor</b>	Evaluar impactos clínicos, revisar casos críticos y aprobar actualizaciones.	Actas de revisión, reportes de incidentes y seguimiento de apelaciones.

### 6.2 Mecanismos de Accountability

#### Documentación

- Model Card describiendo capacidades, limitaciones y contexto de uso del sistema.
- Datasheet de los datos empleados (fuentes, tipos de imágenes, diversidad).
- Ethics statement con análisis de riesgos y principios de mitigación adoptados.

## Supervisión

- Revisión médica humana en todas las decisiones diagnósticas.
- Proceso formal de apelación para usuarios o pacientes en caso de resultados cuestionables.
- Auditoría ética y técnica semestral para revisar desempeño, sesgo y cumplimiento normativo.

## Monitoreo Continuo

- Métricas de *fairness* y desempeño rastreadas automáticamente.
- Alertas de *data drift* (cambio en patrones de datos) o anomalías del modelo.
- Reportes de incidentes generados por el sistema o los usuarios.

## Transparencia

- Comunicación clara y accesible sobre el uso de IA y sus limitaciones.
- Acceso a explicaciones visuales de las decisiones (*XAI – Explainable AI*).

## 6.3 Plan de Respuesta a Incidentes

### Si algo sale mal:

1. **Detección:**  
Identificación automática o manual de errores mediante alertas de rendimiento o reportes de usuarios.
2. **Respuesta:**  
Suspender temporalmente la función afectada y activar revisión médica obligatoria del caso.
3. **Investigación:**  
Analizar registros del modelo (versión, parámetros, entrada) y verificar si hubo sesgo o falla técnica.
4. **Corrección:**  
Ajustar umbrales, retrain del modelo o aplicar actualizaciones de seguridad según el origen del error.
5. **Prevención:**  
Incorporar el caso al repositorio de *lessons learned*, reforzar auditorías y revisar los controles de calidad previos al despliegue.

## Parte 7: Consideraciones de Compliance

### Regulaciones Aplicables

- GDPR (Reglamento General de Protección de Datos – Unión Europea)
- HIPAA (Health Insurance Portability and Accountability Act – EE.UU.)
- Leyes locales de protección de datos personales (Ecuador – Ley Orgánica de Protección de Datos Personales, 2021)
- Leyes anti-discriminación (por el uso de IA en decisiones médicas)

#### 1. GDPR (Unión Europea)

##### ¿Qué requiere?

Protección de datos personales, consentimiento informado, anonimización de información y derecho a la explicación cuando interviene un algoritmo automatizado.

##### ¿Tu proyecto cumple?

Sí, el sistema anonimiza las imágenes ecográficas y no almacena identificadores personales. Además, el modelo es explicable (*Explainable AI*) y mantiene trazabilidad de cada predicción.

##### ¿Qué cambios necesitas?

Implementar un documento formal de consentimiento informado y establecer un registro de auditoría de accesos a los datos.

#### 2. HIPAA (EE.UU.)

##### ¿Qué requiere?

Garantizar la confidencialidad, integridad y disponibilidad de los datos de salud; aplicar medidas de seguridad técnicas (cifrado, autenticación) y administrativas (políticas y controles).

##### ¿Tu proyecto cumple?

Parcialmente. El sistema usa almacenamiento seguro y anonimización, pero requiere incorporar un módulo de cifrado en tránsito y políticas de acceso controlado.

##### ¿Qué cambios necesitas?

Agregar en el *pipeline* del proyecto (en el .ipynb o entorno de despliegue) un sistema de autenticación por usuario y registro de logs de acceso.

### 3. Ley Orgánica de Protección de Datos Personales (Ecuador)

#### ¿Qué requiere?

Obligatoriedad de consentimiento informado, uso legítimo de datos, limitación de finalidad y derecho a rectificación o eliminación de la información.

#### ¿Tu proyecto cumple?

Sí. Los datos utilizados en el entrenamiento provienen de repositorios públicos o previamente anonimizados y se emplean con fines estrictamente académicos y de investigación médica.

#### ¿Qué cambios necesitas?

Incluir una declaración de finalidad en la documentación del modelo indicando que los datos no serán reutilizados fuera del contexto clínico o de investigación.

### 4. Leyes Anti-discriminación (IA médica)

#### ¿Qué requiere?

Evitar sesgos que afecten la igualdad de acceso o diagnóstico entre grupos poblacionales. Se exige demostrar que el modelo ha sido evaluado en distintos grupos y que no produce discriminación indirecta.

#### ¿Tu proyecto cumple?

Sí, ya que el modelo incluye una estrategia de mitigación de sesgos (Parte 5) y monitoreo de métricas de *fairness* para diferentes características demográficas.

#### ¿Qué cambios necesitas?

Mantener auditorías periódicas de equidad y documentar resultados de validaciones cruzadas por género, edad y procedencia de los datos.

## Parte 8: Reflexión y Compromiso

### 8.1 Reflexión Personal

#### 1. ¿Cuál es el dilema ético más difícil de tu proyecto?

El dilema más complejo surge entre la precisión tecnológica y la responsabilidad humana. Si bien el sistema de IA puede superar la capacidad de observación de un médico en la detección de patrones ecográficos, depender completamente de él puede comprometer la autonomía y criterio clínico del profesional.

- **Valores en conflicto:** Eficiencia tecnológica vs. responsabilidad médica.
- **Cómo lo resolverías:** Asegurando que el sistema actúe únicamente como apoyo y no como sustituto, reforzando la revisión médica obligatoria y la trazabilidad de las decisiones.

## 2. Qué aprendiste sobre tu proyecto que no habías considerado?

Aprendí que un sistema de IA no solo debe ser técnicamente sólido, sino también éticamente transparente y socialmente inclusivo. La calidad del dataset y la diversidad de las muestras determinan el grado de justicia y equidad del modelo. También comprendí la necesidad de mantener una comunicación clara sobre los alcances y limitaciones del sistema ante los usuarios médicos.

## 3. Cambiarías algo en tu diseño después de este análisis?

Sí. Incorporaría desde el inicio un módulo de monitoreo ético automatizado, que mida el desempeño y sesgo del modelo en cada iteración, y añadiría un espacio de explicación visual más detallado para los médicos, fortaleciendo la confianza en la herramienta.

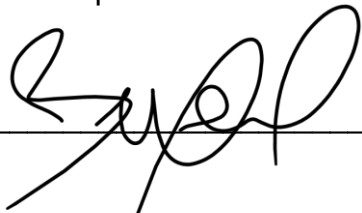
## 8.2 Compromiso Ético

### Declaración de Compromiso

*“Como creadores de este sistema de IA, me comprometo a:”*

- Evaluar métricas de *fairness* y sesgo de manera periódica.
- Mantener la anonimización y seguridad de los datos utilizados.
- Garantizar que el modelo permanezca como apoyo clínico, no sustituto de juicio médico.
- Actuar responsablemente si se identifican errores o daños potenciales.
- Priorizar el bienestar de los usuarios y grupos vulnerables.
- Mantener transparencia sobre las limitaciones y el alcance del sistema.

Firma/Fecha:



Ing. Byron Piedra



Ing. Christian Garcia S., MBA

16/10/2025



### **Bibliografía sugerida (formato APA 7ª edición)**

- European Parliament. (2016). *Reglamento General de Protección de Datos (GDPR)*. Diario Oficial de la Unión Europea.
- U.S. Department of Health and Human Services. (1996). *Health Insurance Portability and Accountability Act (HIPAA)*. Federal Register.
- Asamblea Nacional del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales*. Registro Oficial Suplemento 459.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.
- Zhou, Z.-H. (2021). *Machine learning (2nd ed.)*. Springer.
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.