



PROYECTO INTEGRADOR
Impacto Social y Responsabilidad en el Proyecto de IA

Profesora	GLADYS MARIA VILLEGAS RUGEL
Materia	PROYECTO INTEGRADOR EN INTELIGENCIA ARTIFICIAL
Alumnos	Bolaños Escandón María Fernanda Montaño Cárdenas Fernando Xavier
Fecha	14/10/2025

## Contenido

1.	CONTEXTO DEL PROYECTO.....	1
1.1	Descripción del Sistema.....	1
1.2	Alcance de Implementación .....	1
2.	ANÁLISIS DE STAKEHOLDERS .....	2
3.	EVALUACIÓN DE IMPACTO SOCIAL .....	3
3.1	Impactos Positivos.....	3
3.2	Impactos Negativos y Riesgos .....	3
3.3	Análisis de Equidad.....	4
4.	RIESGOS ÉTICOS ESPECÍFICOS .....	5
4.1	Fairness y Sesgo.....	5
4.2	Privacidad y Protección de Datos .....	5
4.3	Transparencia y Explicabilidad .....	5
4.4	Autonomía y Supervisión Humana .....	5
4.5	Seguridad y Responsabilidad .....	6
5.	ESTRATEGIAS DE MITIGACIÓN.....	7
5.1	Sesgo algorítmico y Fairness.....	7
5.2	Privacidad y Re-identificación .....	7
5.3	Transparencia y Explicabilidad .....	7
5.4	Supervisión Médica y Autonomía.....	7
5.5	Fallos y Accountability .....	8
6.	FRAMEWORK DE RESPONSABILIDAD .....	9
6.1	Cadena de Responsabilidad.....	9
6.2	Mecanismos de Accountability.....	9
6.3	Plan de Respuesta a Incidentes .....	9
7.	COMPLIANCE Y REGULACIONES .....	10
8.	REFLEXIÓN Y COMPROMISO ÉTICO .....	10
8.1	Reflexión Personal y Grupal .....	10
8.2	Compromiso Ético .....	10
	REFERENCIAS .....	1

# IMPACTO SOCIAL Y RESPONSABILIDAD EN EL PROYECTO DE IA

*Autores: María Fernanda Bolaños Escandón y Fernando Xavier Montaña Cárdenas*

Fecha: 13 de octubre de 2025

## 1. CONTEXTO DEL PROYECTO

### 1.1 Descripción del Sistema

**Problema que resuelve:** La descompensación clínica en pacientes con diabetes tipo 2 e hipertensión arterial suele detectarse tarde, aumentando hospitalizaciones y costos. Este proyecto propone un sistema inteligente basado en aprendizaje automático supervisado y series de tiempo para predecir riesgos de descompensación (diabetes, hipertensión o ambas) y emitir alertas tempranas.

**Funcionalidad principal:** El sistema implementa un pipeline modular con:

- Feature engineering clínico (interacciones glucosa-HbA1c, presión arterial, lags, flags)
- Codificación categórica y transformación numérica
- Selección de variables (Chi<sup>2</sup>, ANOVA, MI, LASSO, RFE)
- Modelado supervisado multiclase con output de riesgo por visita (-1=ninguna, 0=diabetes, 1=hipertensión, 2=ambas)

**Datos utilizados:** 300.000 observaciones (100.000 pacientes × 3 visitas), 13 variables clínicas (glucosa, HbA1c, presión arterial, IMC, edad, etc.) con imputación sintética guiada por reglas fisiológicas.

**Usuarios objetivo:**

- Médicos, Instituciones médicas y gestores hospitalarios (alertas y explicaciones interpretables)
- Enfermería (priorización)
- Pacientes (notificaciones futuras)
- Gestores clínicos (monitoreo poblacional)
- Equipos técnicos (mantenimiento del sistema)

**Dominio de aplicación:** Salud digital y medicina de precisión para manejo ambulatorio de pacientes crónicos (diabetes tipo 2, HTA y comorbilidad) en Ecuador. Pensado para integrarse a plataformas de monitoreo y ser interoperable con sistemas clínicos.

### 1.2 Alcance de Implementación

- **Lugar de implementación:** Se propone un piloto inicial con miras a una futura implementación en unidades de atención primaria y hospitales provinciales de Cuenca y Riobamba, donde la prevalencia de diabetes e hipertensión supera el 20%. Actualmente, el desarrollo se encuentra en fase preliminar.
- **Población beneficiada:** En una fase futura, podría beneficiar a más de 15.000 pacientes crónicos en Cuenca y Riobamba, y eventualmente a 120.000 a nivel nacional, si se llegara a implementar y se diera seguimiento. También se proyecta impacto en el trabajo de médicos, enfermeras y gestores clínicos.
- **Toma de decisiones:** El sistema no reemplaza al médico, sino que podría apoyar decisiones clínicas con alertas predictivas, pudiendo facilitar ajustes terapéuticos, priorización de citas y derivaciones oportunas.

## 2. ANÁLISIS DE STAKEHOLDERS

Stakeholder	Descripción	Beneficios	Riesgos / Pérdidas	Poder / Voz
<b>Pacientes con enfermedades crónicas (diabetes tipo 2, hipertensión y/o comorbilidad)</b>	Usuarios finales del sistema; proveen datos fisiológicos (glucosa, presión, IMC, etc.) y reciben alertas predictivas sobre riesgo clínico.	Diagnóstico temprano, reducción de hospitalizaciones, educación preventiva y mayor control sobre la salud.	Riesgo de dependencia tecnológica, ansiedad por falsas alertas y exposición de datos sensibles.	<b>Medio</b> (usuarios finales, pero su voz depende de la representación institucional).
<b>Profesionales de la salud (médicos tratantes, enfermeras, gestores clínicos)</b>	Utilizan el sistema como herramienta de apoyo diagnóstico y priorización de pacientes.	Soporte en la toma de decisiones, detección temprana de descompensaciones y optimización del tiempo clínico.	Riesgo de sobre confianza en el modelo, necesidad de capacitación y adaptación tecnológica.	<b>Alto</b> (actores decisivos en adopción y validación del sistema).
<b>Instituciones médicas y gestores hospitalarios</b>	Administran la infraestructura y los procesos de atención donde se integra el sistema (a futuro).	Mejora de eficiencia operativa, reducción de costos por complicaciones y mejor trazabilidad de pacientes crónicos.	Costos iniciales de integración, resistencia al cambio organizacional y riesgos legales por fallas en decisiones clínicas.	<b>Alto</b> (controlan la adopción institucional y la política de uso).
<b>Desarrolladores e investigadores de IA clínica</b>	Diseñan, entrenan y evalúan modelos predictivos.	Validación científica, avance tecnológico y reputación profesional.	Riesgos éticos si el modelo incurre en sesgos y presión por resultados.	<b>Medio–Alto</b> (control técnico del sistema y su mantenimiento).
<b>Autoridades sanitarias y reguladores (MSP, ARCSA, comités de ética)</b>	Supervisan cumplimiento normativo, privacidad de datos y buenas prácticas clínicas.	Acceso a herramientas basadas en evidencia para planificación de políticas públicas.	Riesgo de falta de comprensión técnica o demora regulatoria ante innovación rápida.	<b>Alto</b> (garantizan legitimidad y cumplimiento legal).

### Consideraciones por Stakeholder

- **Pacientes:** El sistema mejora su calidad de vida mediante alertas y recomendaciones basadas en datos fisiológicos interpretables. Fueron considerados al definir variables comprensibles y rangos clínicamente plausibles. Se garantiza su protección con anonimización ya que no se manejan nombres, solo un ID.
- **Profesionales de la salud e Instituciones médicas:** Apoya su labor clínica al detectar riesgos tempranos y priorizar pacientes críticos. Se incluyeron datos y conceptos clínicos relevantes en el diseño, validando umbrales y relaciones clínicas clave. Su rol se respeta al mantener el sistema como herramienta de apoyo, con trazabilidad y explicabilidad de las decisiones.
- **Desarrolladores e investigadores:** Son responsables del desarrollo ético y técnico del sistema. Participaron en el diseño del pipeline, algoritmos y validaciones. Sus intereses se protegen con buenas prácticas, control de versiones y principios de IA confiable [1].

### 3. EVALUACIÓN DE IMPACTO SOCIAL

El sistema busca fortalecer la atención de pacientes con diabetes tipo 2, hipertensión o ambas, aplicando modelos supervisados y análisis de series de tiempo. Su impacto social combina beneficios clínicos y de gestión con riesgos éticos que deben mitigarse mediante el diseño responsable y gobernanza algorítmica.

#### 3.1 Impactos Positivos

- **Prevención y calidad de vida de los pacientes:** El modelo predice eventos críticos a partir de patrones fisiológicos (glucosa, HbA1c, presión arterial e IMC), reduciendo hospitalizaciones evitables y mejorando el seguimiento ambulatorio.
  - *Beneficiarios:* pacientes crónicos y médicos tratantes.  
*Evidencia:* macro F1 85.29 % y accuracy multiclase de 96.80 % en validaciones con datos clínico-sintéticos.
- **Optimización de recursos sanitarios:** Permite priorizar pacientes de alto riesgo y planificar la atención hospitalaria con mayor eficiencia.
  - *Beneficiarios:* hospitales públicos y privados y aseguradoras.  
*Evidencia:* las alertas predictivas podrían ayudar a reducir costos y mejorar la gestión preventiva.
- **Reducción de inequidades:** La integración con telemedicina puede extender la cobertura sanitaria a zonas rurales, promoviendo equidad en el acceso.
  - *Beneficiarios:* comunidades con baja disponibilidad médica.
  - *Evidencia:* su diseño modular permite la visualización de datos y alertas desde distintos contextos.

#### 3.2 Impactos Negativos y Riesgos

- **Sesgos algorítmicos:** Riesgo de sub-representar grupos demográficos (mujeres y adultos mayores).
  - *Severidad:* alta
  - *Probabilidad:* media.
  - *Mitigación:* Usa estratificación por edad y género en el split/entrenamiento + class\_weight para compensar subrepresentación; evaluar métricas por subgrupo y ajustar umbrales del clasificador si detectas brechas.
- **Privacidad y seguridad:** Posible exposición de datos sensibles.
  - *Severidad:* alta
  - *Probabilidad:* baja.
  - *Mitigación:* anonimización ya que se manejan solo ID de pacientes no datos sensibles como nombres o direcciones.
- **Dependencia tecnológica:** Uso excesivo sin interpretación médica.
  - *Severidad:* media-alta
  - *Probabilidad:* media.
  - *Mitigación:* explicabilidad (SHAP, LIME) y supervisión clínica obligatoria.

### **3.3 Análisis de Equidad**

El sistema reduce desigualdades estructurales al priorizar prevención y cobertura remota, aunque el acceso inicial beneficiará más a entornos urbanos. Los riesgos recaen sobre los desarrolladores y pacientes, por lo que se propone auditoría médica independiente, inclusión de datos diversos y capacitación digital.

En conjunto, su impacto social es positivo, éticamente justificable y clínicamente relevante, contribuyendo a un ecosistema sanitario más preventivo, equitativo y eficiente.

## 4. RIESGOS ÉTICOS ESPECÍFICOS

El desarrollo de sistemas inteligentes en salud exige un marco ético riguroso, pues sus decisiones pueden afectar directamente la vida y bienestar de los pacientes. En este proyecto se identifican los principales riesgos éticos asociados a la implementación de un modelo predictivo de descompensaciones clínicas basado en IA y series de tiempo.

### 4.1 Fairness y Sesgo

**Riesgo:** discriminación por género, edad o procedencia geográfica.

**Descripción:** la menor representación de mujeres, adultos mayores y pacientes rurales en los datos clínicos y la falta de datos balanceados puede inducir sesgos de predicción.

- *Severidad:* Alta
- *Grupo afectado:* poblaciones vulnerables.
- *Mitigación:* evaluación periódica de métricas de *fairness* y reentrenamiento con datos balanceados, siguiendo las recomendaciones de modelos éticos propuestos por la IEEE Global Initiative [2].

### 4.2 Privacidad y Protección de Datos

**Riesgo:** exposición o re-identificación de información médica sensible.

**Descripción:** el sistema maneja variables como glucosa y presión arterial, que requieren medidas de anonimización estricta.

- *Severidad:* Alta
- *Grupo afectado:* pacientes.
- *Mitigación:* cifrado de extremo a extremo y cumplimiento con la Ley Orgánica de Protección de Datos Personales de Ecuador, en concordancia con las directrices ISO/IEC 27701 [3].

### 4.3 Transparencia y Explicabilidad

**Riesgo:** falta de comprensión del funcionamiento del modelo.

**Descripción:** el uso de algoritmos complejos como Random Forest y LSTM, podría dificultar la interpretación de resultados clínicos.

- *Severidad:* Media
- *Grupo afectado:* pacientes y médicos tratantes.
- *Mitigación:* uso de técnicas de explicabilidad como SHAP y LIME, que permiten visualizar la contribución de cada variable [4]. Además, se puede crear un manual de usuario.

### 4.4 Autonomía y Supervisión Humana

**Riesgo:** dependencia excesiva de notificaciones automáticas.

**Descripción:** el personal médico podría confiar plenamente en las predicciones sin validar clínicamente los resultados.

- *Severidad:* Media- Alta

- *Grupo afectado:* pacientes y profesionales de la salud.  
*Mitigación:* mantener supervisión médica obligatoria, auditorías clínicas y capacitación en uso ético de IA.

## 4.5 Seguridad y Responsabilidad

**Riesgo:** fallos de predicción con consecuencias clínicas.

**Descripción:** errores del modelo podrían generar decisiones inadecuadas de tratamiento.

- *Severidad:* Alta
- *Grupo afectado:* pacientes y hospitales.
- *Mitigación:* pruebas de validación continua, registro de decisiones y protocolos de respuesta ante errores, promoviendo accountability institucional.

\*Los responsables directos son los desarrolladores que deben mantener una supervisión constante con profesionales de la salud y auditorías constantes [5].



## 5. ESTRATEGIAS DE MITIGACIÓN

### 5.1 Sesgo algorítmico y Fairness

**Estrategia:** Reentrenar el modelo con más datos (series de tiempo) y aplicar métricas de *equal opportunity* por género y edad para evitar discriminación.

**Tipo:** Técnica.

**Implementación:**

1. Evaluar sesgos con *Fairlearn*.
2. Ajustar pesos de clase y umbrales de decisión.
3. Monitorear métricas de equidad periódicamente.

**Responsable:** Desarrolladores y Equipo de ciencia de datos.

**Efectividad:** Alta.

### 5.2 Privacidad y Re-identificación

**Estrategia:** Anonimizar identificadores y cifrar datos sensibles conforme a la Ley Orgánica de Protección de Datos Personales [6].

**Tipo:** Técnica / Política.

**Implementación:**

1. Hash de patient\_id y cifrado TLS.
2. Auditorías periódicas de seguridad.

**Responsable:** Desarrolladores y Equipo de ciencia de datos.

**Efectividad:** Alta.

### 5.3 Transparencia y Explicabilidad

**Estrategia:** Incorporar reportes interpretables con SHAP y documentación técnica visible para médicos.

**Tipo:** Técnica / Diseño.

**Implementación:**

1. Visualizar impacto de variables.
2. Explicar límites y confianza del modelo.

**Responsable:** Desarrolladores y Equipo de ciencia de datos.

**Efectividad:** Media-Alta.

### 5.4 Supervisión Médica y Autonomía

**Estrategia:** Mantener revisión clínica obligatoria para toda alerta automática y capacitar al personal.

**Tipo:** Educación / Política.

**Implementación:**

1. Validación doble (IA + médico)
2. Talleres de uso responsable.

**Responsable:** Comité clínico y científicos de datos.

**Efectividad:** Media.

## 5.5 Fallos y Accountability

**Estrategia:** Establecer trazabilidad completa de decisiones y protocolo ante errores de predicción.

**Tipo:** Política / Técnica.

**Implementación:**

1. Registro de logs automáticos.
2. Revisión médica de casos fallidos.

**Responsable:** Desarrolladores, Comité clínico y científicos de datos.

**Efectividad:** Alta.

Estas medidas combinan acciones técnicas, médicas y éticas para garantizar equidad, privacidad y seguridad en el uso clínico del sistema, alineadas con las directrices de la IEEE sobre IA responsable [4].

## 6. FRAMEWORK DE RESPONSABILIDAD

### 6.1 Cadena de Responsabilidad

Rol	Responsabilidades	Rendición de Cuentas
<b>Desarrolladores</b>	Implementar código limpio, reproducible y seguro; mantener la integridad del pipeline de preprocesamiento y del modelo predictivo.	<i>Code reviews</i> semanales, control de versiones en GitHub, y testing automatizado.
<b>Data Scientists</b>	Curar y validar los datos clínicos, aplicar imputación médico-realista, controlar sesgos y monitorear métricas de fairness y drift.	Reportes mensuales de calidad de datos y auditorías internas de sesgo.
<b>Product Owner / Coordinador Clínico</b>	Definir requerimientos funcionales, validar interpretabilidad de resultados y priorizar decisiones clínicas sobre las automáticas.	Revisión semanal del backlog y reportes de cumplimiento clínico.
<b>Organización (Institución de Salud / Proyecto IA)</b>	Establecer políticas éticas, asignar recursos para supervisión y garantizar cumplimiento con normas de protección de datos.	Informes trimestrales a la dirección médica y comité de ética institucional.

### 6.2 Mecanismos de Accountability

#### Documentación:

- *Model Card* detallando propósito, métricas (macro F1 85.29 % y accuracy multiclase de 96.80 %), y limitaciones del modelo.
- *Datasheet for Datasets* documentando el origen (dataset clínico semi-sintético), proceso de imputación y variables sensibles.
- *Ethics Statement* con evaluación de sesgos, privacidad y medidas de mitigación.

#### Supervisión:

- Revisión humana obligatoria en alertas críticas de descompensación.
- Canal de apelación médica en caso de predicciones erróneas.
- Auditorías de desempeño y sesgo cada tres meses.

#### Monitoreo Continuo:

- Seguimiento de métricas de fairness y precisión en tiempo real.
- Detección de *drift* en distribuciones clínicas (glucosa, presión diastólica y sistólica).
- Registro y análisis de incidentes mediante logs automatizados.

#### Transparencia:

- Comunicación clara a los médicos sobre el rol asistencial del sistema.
- Explicación visual de cada predicción (usando SHAP/LIME).
- Acceso a documentación y reportes para validación externa.

### 6.3 Plan de Respuesta a Incidentes

1. **Detección:** Alertas automáticas ante anomalías en métricas de performance o fairness.
2. **Respuesta inmediata:** Suspensión temporal del modelo y revisión por el equipo clínico.
3. **Investigación:** Análisis de logs y revisión de la trazabilidad completa de la decisión.
4. **Corrección:** Reentrenamiento del modelo o ajuste de umbrales si se confirma error sistemático.
5. **Prevención:** Actualización de políticas, mejora de validaciones y auditorías preventivas.

Este marco garantiza una cadena de responsabilidad clara, trazabilidad técnica y ética, y un ciclo continuo de supervisión y mejora. Con ello, el sistema mantiene seguridad, confiabilidad y transparencia en su aplicación clínica.

## 7. COMPLIANCE Y REGULACIONES

El sistema inteligente de predicción clínica opera en el dominio de **salud digital**, por lo que debe alinearse con normativas de protección de datos, ética médica y uso responsable de inteligencia artificial.

### Regulaciones aplicables:

- **Ley Orgánica de Protección de Datos Personales:** exige consentimiento informado, anonimización y uso legítimo de datos sensibles.
- **Cumplimiento:** total; el sistema anonimiza identificadores y cifra la información clínica en tránsito y reposo.
- **Reglamento General de Protección de Datos (GDPR - UE):** referencia internacional adoptada para buenas prácticas de privacidad y derecho al olvido. **Cumplimiento:** parcial; se proyecta incorporar mecanismos de eliminación bajo solicitud del paciente.
- **AI Act (Unión Europea, 2024 – categoría de alto riesgo):** los sistemas médicos predictivos deben garantizar trazabilidad, explicabilidad y supervisión humana. **Cumplimiento:** en progreso; el pipeline incluye *model cards*, auditorías de sesgo y revisión médica obligatoria.
- **Ley de Derechos del Paciente (Ecuador, 2015):** garantiza autonomía, confidencialidad y acceso a la información médica.
- **Cumplimiento:** completo; los resultados del sistema se interpretan bajo criterios clínicos.

## 8. REFLEXIÓN Y COMPROMISO ÉTICO

### 8.1 Reflexión Personal y Grupal

El principal dilema ético del proyecto es equilibrar la precisión algorítmica con la autonomía médica, priorizando siempre la supervisión profesional. El análisis evidenció el impacto social de la IA clínica y la necesidad de integrar equidad, explicabilidad y métricas éticas desde el diseño. Se propone incluir monitoreo automatizado de sesgos y auditorías éticas externas para garantizar trazabilidad y confianza.

### 8.2 Compromiso Ético

María Fernanda Bolaños Escandón y Fernando Xavier Montaña Cárdenas, como creadores de este sistema de inteligencia artificial para monitoreo clínico, nos comprometemos a:

- Diseñar, implementar y mantener sistemas de inteligencia artificial que promuevan el bienestar humano, la equidad y la transparencia.
- Evaluar periódicamente la precisión, robustez y equidad del modelo para evitar sesgos o impactos negativos.
- Garantizar la protección, anonimización y uso responsable de los datos sensibles.
- Mantener la supervisión humana en decisiones críticas y la comunicación clara de las limitaciones del sistema.
- Actuaremos de forma responsable, ética y proactiva ante cualquier incidente o riesgo identificado.
- Fomentaremos la mejora continua y la rendición de cuentas en todas las etapas del ciclo de vida del proyecto.

Firmas de responsabilidad:

## REFERENCIAS

- [1] European Commission, *Ethics Guidelines for Trustworthy AI*, Brussels, Belgium: European Commission, 2021. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [2] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, IEEE Standards Association, 2019.
- [3] ISO/IEC 27701:2019, *Security techniques — Extension to ISO/IEC 27001 and 27002 for privacy information management*, ISO, 2019.
- [4] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems (NIPS)*, vol. 30, pp. 4765–4774, 2017.
- [5] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, IEEE Standards Association, 2019.
- [6] República del Ecuador, Ley Orgánica de Protección de Datos Personales, Registro Oficial Suplemento 459, 26 mayo 2021. [En línea]. Disponible en: <https://www.registroficial.gob.ec>