



# Tecnológico de Monterrey

**Inteligencia Artificial Avanzada para la Ciencia de Datos**

**Reto**

**Equipo 1:**

Samuel Pelaez Aleman - A01026093

9 de septiembre del 2024

## **Normativa Asociada al Uso de Datos**

El dataset utilizado contiene información sobre pacientes, como edad, sexo, presión arterial y niveles de colesterol, que podría clasificarse como datos sensibles de salud. Aunque el dataset no incluye información personal identificable (PII), es fundamental considerar las normativas que regulan el manejo de estos datos, especialmente en contextos de salud.

### HIPAA (Health Insurance Portability and Accountability Act - EE.UU.)

La HIPAA establece estándares para la protección de información médica sensible, conocida como PHI (Protected Health Information). Los datos utilizados en este proyecto, al estar anonimizados, no incluyen información que permita identificar directamente a las personas, por lo que se cumple con las regulaciones de HIPAA. Además, el uso de datos anonimizados minimiza el riesgo de violaciones de privacidad, conforme a los estándares de la normativa (HHS, 2022).

### GDPR (General Data Protection Regulation - UE)

El Reglamento General de Protección de Datos (GDPR) de la Unión Europea también establece estándares estrictos para el manejo de datos personales, especialmente los datos de salud. Aunque este proyecto no se desarrolla en Europa, los principios de GDPR son relevantes cuando se manejan datos sensibles a nivel global. Al usar datos anonimizados, se garantiza que no se incurre en violaciones de GDPR (GDPR, 2024).

### Política de Kaggle

Kaggle, la plataforma desde la que se obtuvo el dataset, impone estrictas políticas de uso de datos. Los usuarios de la plataforma deben asegurar que los datos subidos no contienen información personal sin el consentimiento adecuado y cumplen con las leyes aplicables, como HIPAA y GDPR. Esto significa que el uso del dataset para este proyecto es legal y está alineado con las políticas de la plataforma (Kaggle, 2024).

## **Garantías de Cumplimiento Normativo**

### Anonimización de los datos

El dataset utilizado no incluye PII (Personal Identifiable Information), lo que reduce significativamente los riesgos de incumplimiento normativo.

Uso seguro de los datos: El proyecto ha tomado medidas para asegurar que los datos se mantengan seguros durante el proceso de entrenamiento y evaluación del modelo.

Minimización de datos: Solo se utilizaron las variables necesarias para el análisis (edad, sexo, presión arterial, colesterol, etc.), asegurando un uso responsable y limitado de los datos.

## **Cumplimiento Ético en la Herramienta**

El modelo entrenado para la clasificación de medicamentos fue diseñado con un enfoque ético que minimiza el riesgo de sesgos. Dado que la información utilizada no contiene identificadores personales, los principales riesgos éticos están relacionados con posibles sesgos implícitos en el dataset o el modelo.

### Prevención de Sesgos

Equidad: Se debe asegurar que el modelo no discrimine a ciertos grupos de pacientes por características como edad, sexo o condiciones preexistentes. Para ello, es importante evaluar el rendimiento del modelo en subgrupos de datos diversos.

Transparencia: Los algoritmos deben ser interpretables, de modo que las decisiones del modelo puedan ser auditadas y corregidas en caso de detectar posibles sesgos.

## **Riesgos Éticos y Escenarios de Mal Uso**

Si bien la herramienta está diseñada para operar de forma ética, su mal uso o negligencia podría derivar en situaciones problemáticas:

### Discriminación de Pacientes

Si el modelo se entrena con un dataset que contiene sesgos no detectados, puede haber un riesgo de discriminación en la clasificación de medicamentos para ciertos grupos demográficos. Es crucial monitorear la equidad del modelo y realizar auditorías periódicas para garantizar un trato justo.

### Uso Incorrecto en Decisiones Médicas

El modelo está destinado a ser una herramienta de apoyo para la toma de decisiones, no un reemplazo de la supervisión médica. Si se utiliza sin revisión adecuada por profesionales de la salud, las decisiones incorrectas podrían poner en riesgo a los pacientes.

### Manipulación de Datos

Si se modifican los datos de entrada de manera maliciosa, el modelo podría clasificar erróneamente los medicamentos, afectando negativamente la integridad del sistema y el bienestar de los pacientes.

## **Conclusión**

Este proyecto cumple con las normativas clave para el manejo de datos de salud al utilizar datos anonimizados y seguir buenas prácticas de seguridad y minimización de datos.

Además, se han identificado y mitigado los riesgos éticos más relevantes, como la discriminación y el uso incorrecto de la herramienta.