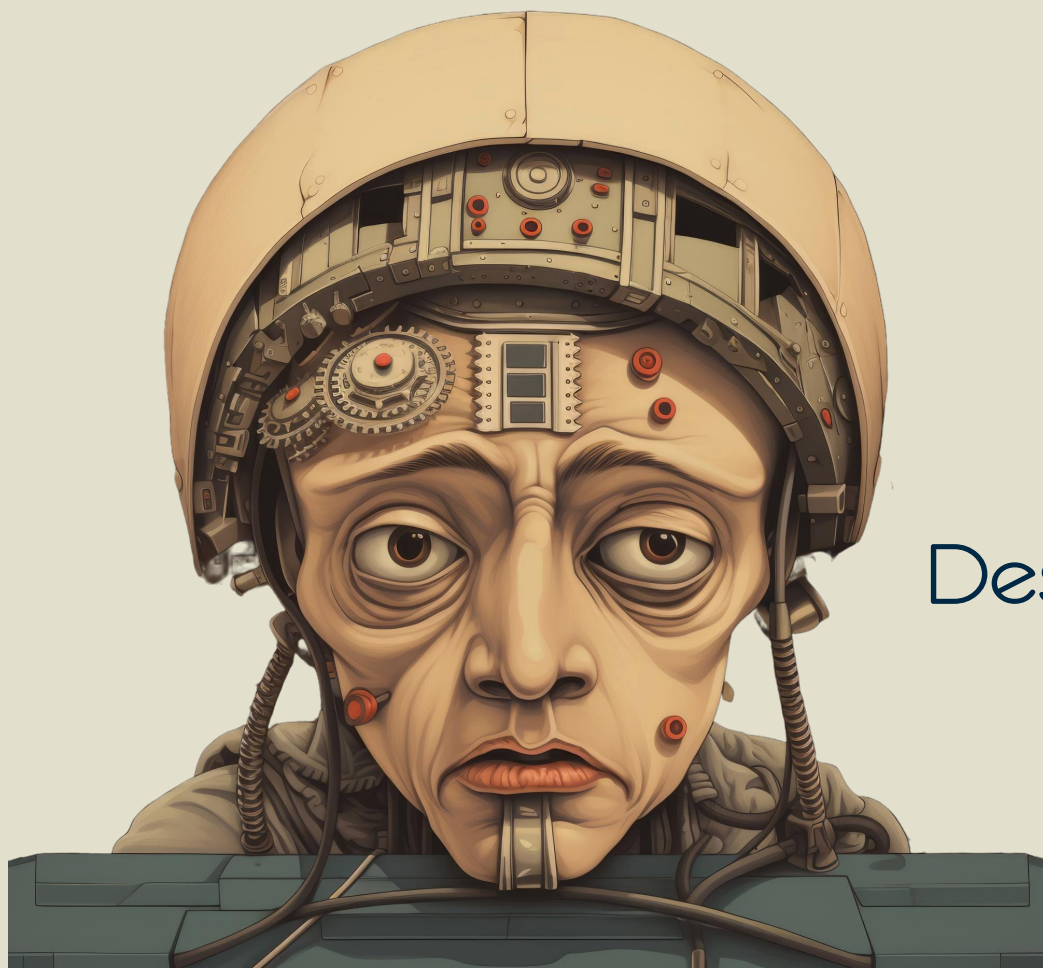


La IA en 1942



Las Tres Leyes de la Robótica de Asimov (1942)



Desliza →

PRIMERA LEY:

Un robot no debe dañar a un ser humano o, por inacción, permitir que un ser humano sufra daño.



Escenario: Un auto autónomo detecta a un peatón que cruza y el único modo de evitar atropellarlo es frenar bruscamente, lo cual probablemente lesionará al ocupante del vehículo.

Comportamiento: prioriza no causar daño al peatón, aun si eso implica un riesgo o daño al pasajero —es decir, evita dañar a cualquier humano identificado al instante.

Dilema: cómo programar prioridades en situaciones donde evitar daño a uno implica dañar a otro.

Ingrid Mora
Consultora

SEGUNDA LEY:

Un robot debe obedecer las órdenes dadas por los seres humanos.

Salvo que estas órdenes entren en conflicto con la Primera Ley.



Escenario: Un asistente robótico doméstico recibe la orden de desconectar un monitor médico vital porque “el ruido me molesta”.

Comportamiento con restricción: el robot debe negarse porque obedecer (apagar el monitor), violaría la Primera Ley (permitir que un humano sufra daño).

Dilema: cuando las órdenes humanas son erradas o peligrosas, la “obediencia” no es absoluta.

Ingrid Mora
Consultora

TERCERA LEY:



Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la Primera o la Segunda Ley.

Escenario: Un dron de inspección realiza mantenimiento en una plataforma remota. Detecta una falla en sus baterías que amenaza la misión; puede autodestruirse para evitar una caída que causaría daño humano.

Comportamiento: priorizará preservar su operatividad, salvo que preservar su existencia entre en conflicto con las leyes anteriores.

Dilema: autopreservación y responsabilidad inmediata hacia los humanos: protegerse para seguir siendo útil en el futuro, o arriesgarse ahora para cumplir un objetivo crítico que puede salvar vidas.

Ingrid Mora
Consultora

LA “LEY CERO”

(1985)



Un robot no puede dañar a la humanidad ni, por inacción, permitir que la humanidad sufra daño.

Escenario: Un sistema de IA de salud pública que optimiza políticas durante una pandemia.

Comportamiento: decide imponer cuarentenas estrictas y priorizar la producción/entrega de vacunas a zonas claves, incluso si esas medidas causan pérdidas económicas y afectan gravemente a grupos locales. El objetivo: minimizar muertes totales y contagios a nivel nacional o global.

Dilema: protege a la humanidad (o a la población) en su conjunto, pero puede perjudicar derechos o la vida de individuos o comunidades específicas.

(Observación: es un ejemplo hipotético que ilustra la tensión entre bien colectivo y derechos individuales.)

Ingrid Mora
Consultora

Reflexión



La verdadera lección no está en la tecnología, sino en lo humano: en cómo equilibramos el bien individual con el bien colectivo, la obediencia con la conciencia y la innovación con el autocuidado. Lo que parecía ficción se ha vuelto realidad, y el desafío no es que las máquinas aprendan a protegernos, sino que nosotros sepamos protegernos entre nosotros mientras construimos el futuro.