



# **Inteligencia Artificial en medicina: ¿Herramienta o cambio de paradigma?**

**Dr. Cristián González**

**Comisión Salud Digital COLMED**

**Abril 2026**

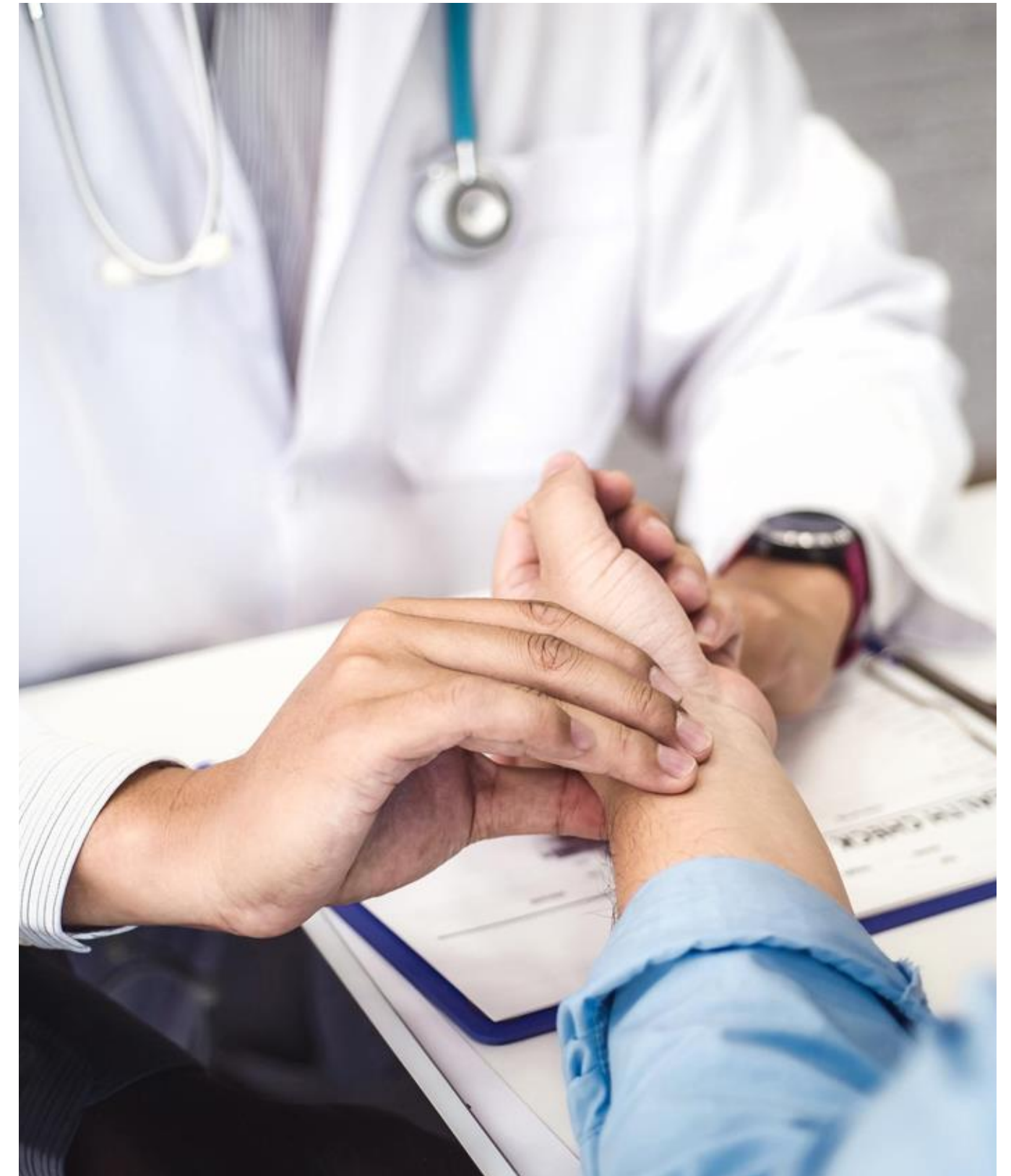


**COLEGIO MÉDICO DE CHILE**



# Temas

- **Introducción: La curva de ascenso de la IA.**
- **IA como herramienta para la medicina.**
- **El cambio de paradigma.**
- **El rol de los profesionales en la nueva era.**
- **Ideas finales.**



# ¿Qué es la IA? Definiciones oficiales OMS



- Mythos de Anthropic es la primera vez que una empresa de IA **frenó voluntariamente el lanzamiento de un modelo** porque el modelo era demasiado capaz de encontrar vulnerabilidades en software.
- Identificó miles de ellas, incluyendo en todos los sistemas operativos y navegadores principales.
- OpenAI en 2019: retrasaron un modelo porque temían que generara "lenguaje engañoso y abusivo a escala" y luego lanzaron modelos que hicieron exactamente eso. **La presión comercial y la competencia entre laboratorios** hace casi imposible que cualquier actor individual frene el proceso, aunque quiera.





# Panorama general

CONCENTRACIÓN DE VALOR — LAS 5 EMPRESAS MÁS GRANDES DEL MUNDO SON COMPAÑÍAS DE IA O HABILITADAS POR IA (ABRIL 2026)

**\$4,5T**

capitalización de Nvidia — primera empresa en superar \$4 billones en la historia

**7,0%**

del S&P500 es solo Nvidia — más que todo el sector energía o utilities de EEUU

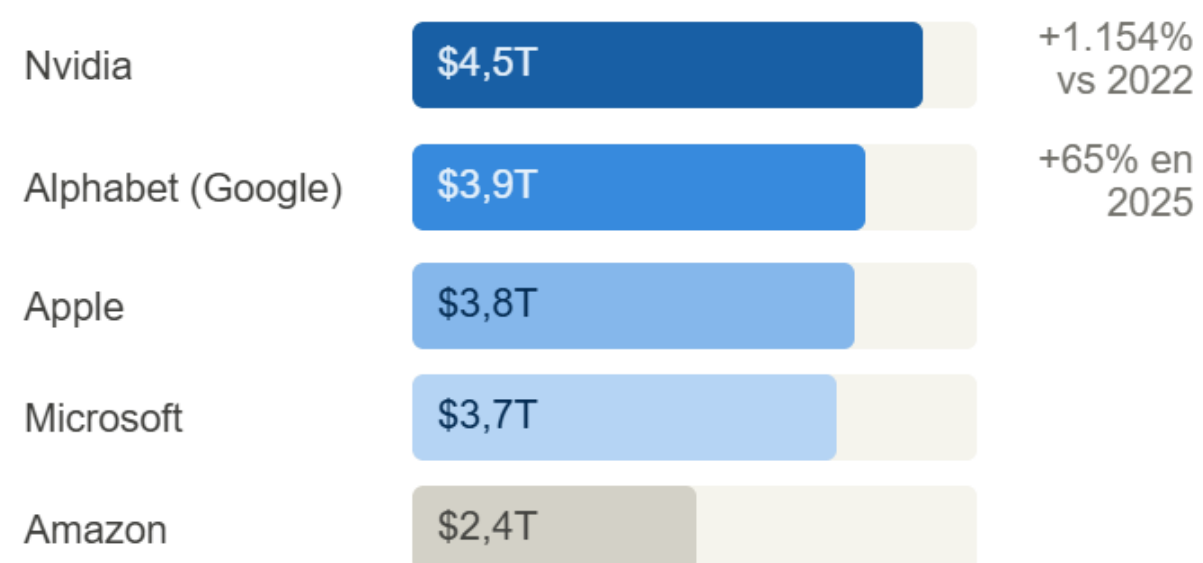
**18%**

del S&P500 concentrado en solo 3 empresas: Nvidia, Apple y Microsoft

**\$67B**

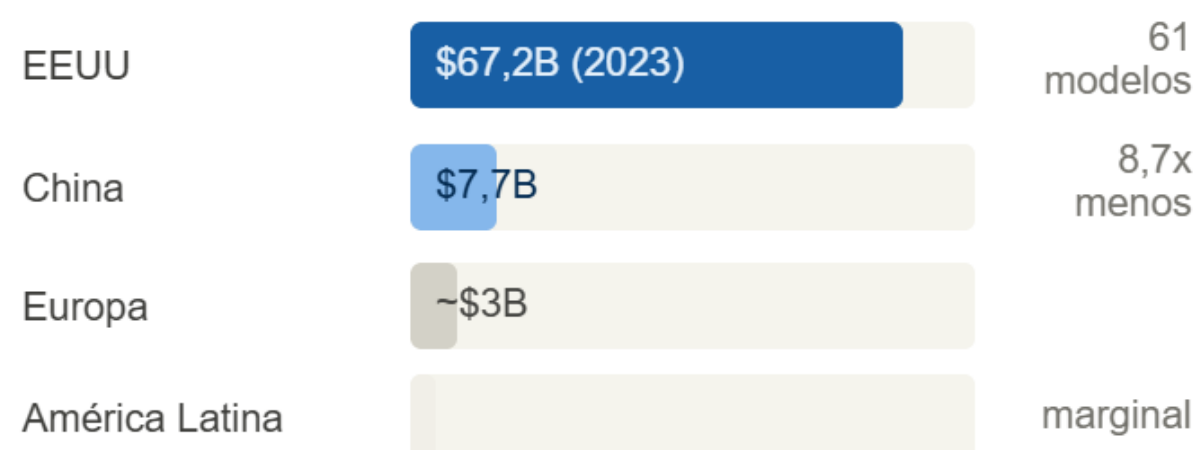
inversión privada en IA captada por EEUU en 2023 — 8,7 veces más que China

## CAPITALIZACIÓN DE MERCADO — GRANDES 5 DE IA (2025–2026)



Concentración S&P500 en máximos históricos según Bloomberg/Apollo (sept. 2025)

## INVERSIÓN PRIVADA GLOBAL EN IA — CONCENTRACIÓN GEOGRÁFICA



**Nvidia: generador de riqueza concentrado**

\$2M+ de ganancia neta por empleado al año. Convierte más del 50% de sus ingresos en ganancia after-tax. Empleados: ~36.000. Capitalización: \$4,5 billones.





# Panorama general

## DESPLAZAMIENTO LABORAL — DATOS GLOBALES Y DE EEUU

40%

de empleos globales con exposición significativa a IA (60% en países de alto ingreso) · FMI 2024

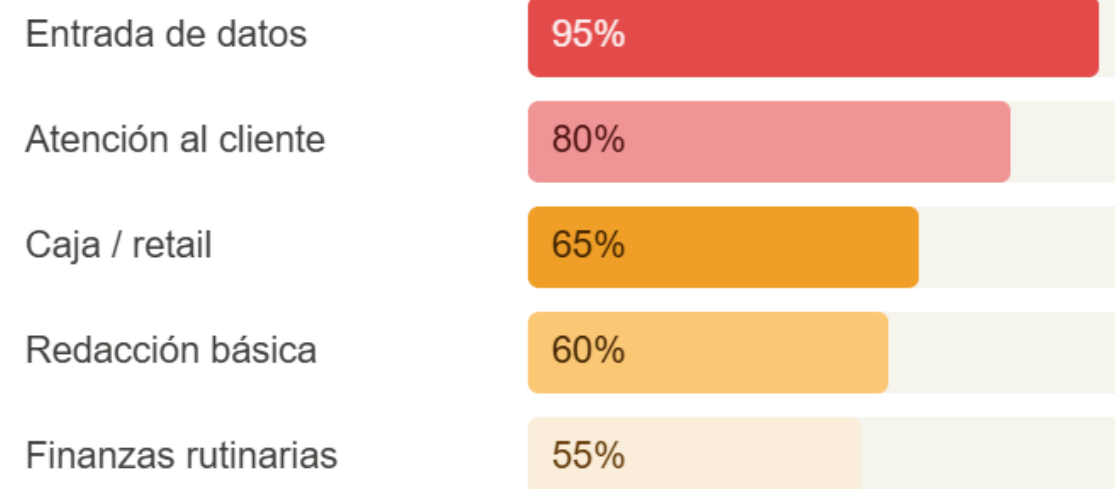
92M

de puestos proyectados a desaparecer para 2030 · WEF Future of Jobs 2025 · (pero 170M nuevos)

41%

de empleadores planea reducir su fuerza laboral para 2030 por IA · WEF 2025

### EMPLEOS EN EEUU CON MAYOR RIESGO DE AUTOMATIZACIÓN



### LA PARADOJA DE LA CLASE MEDIA — EL VACIAMIENTO DEL MEDIO

La IA no ataca primero los trabajos manuales (requieren destreza física) ni los altamente creativos. Ataca los **trabajos cognitivos de rutina de clase media**: administrativos, análisis estructurado, documentación, atención al cliente.

Hallazgo IMF 2025: empleos en zonas con alta demanda de habilidades IA tienen **3,6% menos empleo total** a 5 años que zonas con menor adopción. Los empleos de entrada han caído **15% interanual** en EEUU.

**Dato crítico de género:** 79% de mujeres empleadas en EEUU trabajan en empleos con alto riesgo de automatización, vs 58% de hombres.



# ¿Qué es la IA? Definiciones oficiales OMS



El informe de OMS Europa (2025) adopta las definiciones del documento madre Ethics and Governance of AI for Health (OMS, 2021):

- **Inteligencia artificial:** rama de las ciencias de la computación, la estadística y la ingeniería que utiliza algoritmos o modelos para realizar tareas y exhibir comportamientos como aprender, tomar decisiones y hacer predicciones.
- **Sistema de IA:** sistema basado en máquinas que puede, para un conjunto de objetivos definidos por humanos, hacer predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con distintos niveles de autonomía.
- **Machine learning** (aprendizaje automático): subconjunto de la IA que permite a los algoritmos informáticos aprender a través de datos, sin ser programados explícitamente, para realizar una tarea. Los llamados grandes modelos de lenguaje son un subconjunto adicional del machine learning, entrenados en grandes cantidades de datos textuales para comprender, generar y responder al lenguaje humano



# Los 6 principios éticos de la OMS para la IA en salud



Fuente: Ethics and Governance of Artificial Intelligence for Health · OMS 2021





# ¿Para qué se usa en salud?

Según la encuesta a 50 países de la OMS Europa (2024-2025), las aplicaciones más extendidas son

- El diagnóstico asistido por IA (en radiología, dermatología u oftalmología), utilizado por el 64% de los estados miembros, con el 30% considerándolo ya establecido. La segunda aplicación más frecuente fueron las plataformas conversacionales (chatbots) para asistencia al paciente, empleadas por el 50% de los estados miembros.
- Las principales motivaciones declaradas por los países para adoptar IA fueron: mejorar la atención al paciente y los resultados de salud (98% de los estados miembros), reducir la presión sobre el personal sanitario (92%) y aumentar la eficiencia del sistema de salud (90%).





# La curva de ascenso

**1.451**

dispositivos IA aprobados por la FDA  
acumulados a dic. 2025  
vs. 6 en 2015

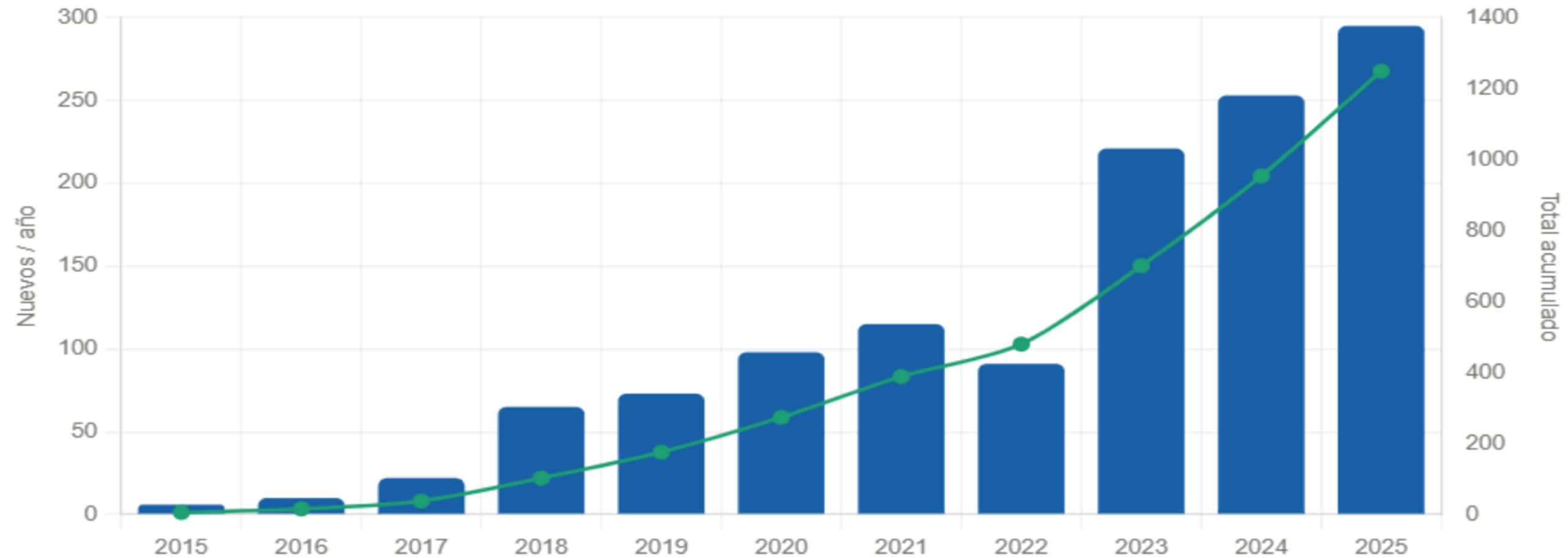
**\$375B**

proyección del mercado de IA en  
medicina para 2032  
desde \$22B en 2024 · CAGR 42%

**66%**

de médicos en EE.UU. ya usan IA en  
su práctica (2024)  
vs. 38% en 2023

■ Aprobaciones FDA anuales    ■ Acumulado total



Fuentes: FDA AI/ML-Enabled Medical Devices List · IntuitionLabs 2025 · Credence Research 2026 · AMA Survey 2024

# Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial.



Ensayo clínico aleatorizado · JAMA Network Open · Oct 2024

## LLM Influence on Diagnostic Reasoning

Goh E, Gallo R et al. — Stanford · Harvard / Beth Israel · Microsoft Research  
DOI: 10.1001/jamanetworkopen.2024.40969

### Diseño

RCT multicéntrico remoto  
4–6 viñetas clínicas/médico

### Participantes

Medicina interna, familiar  
y emergencias

### 3 grupos

Control · Médico+LLM  
LLM solo (GPT-4)

Puntaje de razonamiento diagnóstico por caso (mediana)

GPT-4 solo

92%

P=0.03 ✓

Médico + GPT-4

76%

P=0.60 X

Médico solo (control)

74%

referencia

### Hallazgo clave

Diferencia médico+LLM vs. control: solo 2 puntos — NO significativa.  
Tiempo: 519 s con LLM vs. 565 s sin IA. GPT-4 solo: +16 puntos vs. control (P=0,03).

### Interpretación de los autores

El acceso al LLM no basta: los médicos lo usaron como motor de búsqueda, no como herramienta de razonamiento integral. Se necesita nuevo desarrollo en interacción humano-IA para realizar su potencial clínico.

Goh E et al. JAMA Netw Open. 2024;7(10):e2440969 · Open Access CC-BY





# IA como herramienta

## Diagnostico e imagen medica

Reconocimiento de patrones en imágenes clínicas. La especialidad con mayor volumen de evidencia y aprobaciones regulatorias.

*Radiología · dermatología · patología digital · oftalmología · endoscopia · cardio-imagen*

## Predicción de riesgo y alertas

Algoritmos que anticipan eventos adversos antes de que sean clínicamente evidentes, usando datos de HCE y señales fisiológicas.

*Sepsis · deterioro UCI · readmisión · riesgo cardiovascular · detección temprana cáncer*

## Personalización del tratamiento

Sistemas que ajustan intervenciones a las características individuales del paciente: perfil genómico, comorbilidades, respuesta previa.

*Dosificación insulina · planificación radioterapia · estratificación depresión · oncología de precisión*

## Documentación y administración

Automatización de tareas repetitivas que consumen tiempo clínico sin agregar valor directo al paciente.

*Notas clínicas por voz · codificación diagnóstica · triaje · agendamiento · resumen de alta*

## Paciente y acceso

Experiencia del paciente y acceso

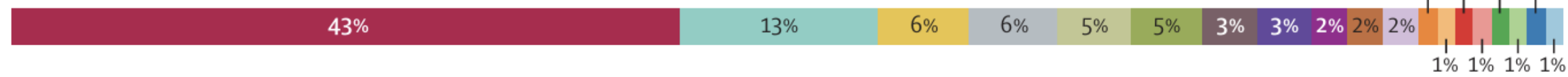
Herramientas que empoderan al paciente, mejoran la comunicación clínica y amplían el acceso en zonas con escasez de especialistas.

*Chatbots de síntomas · adherencia terapéutica · traducción de informes · telemedicina guiada*

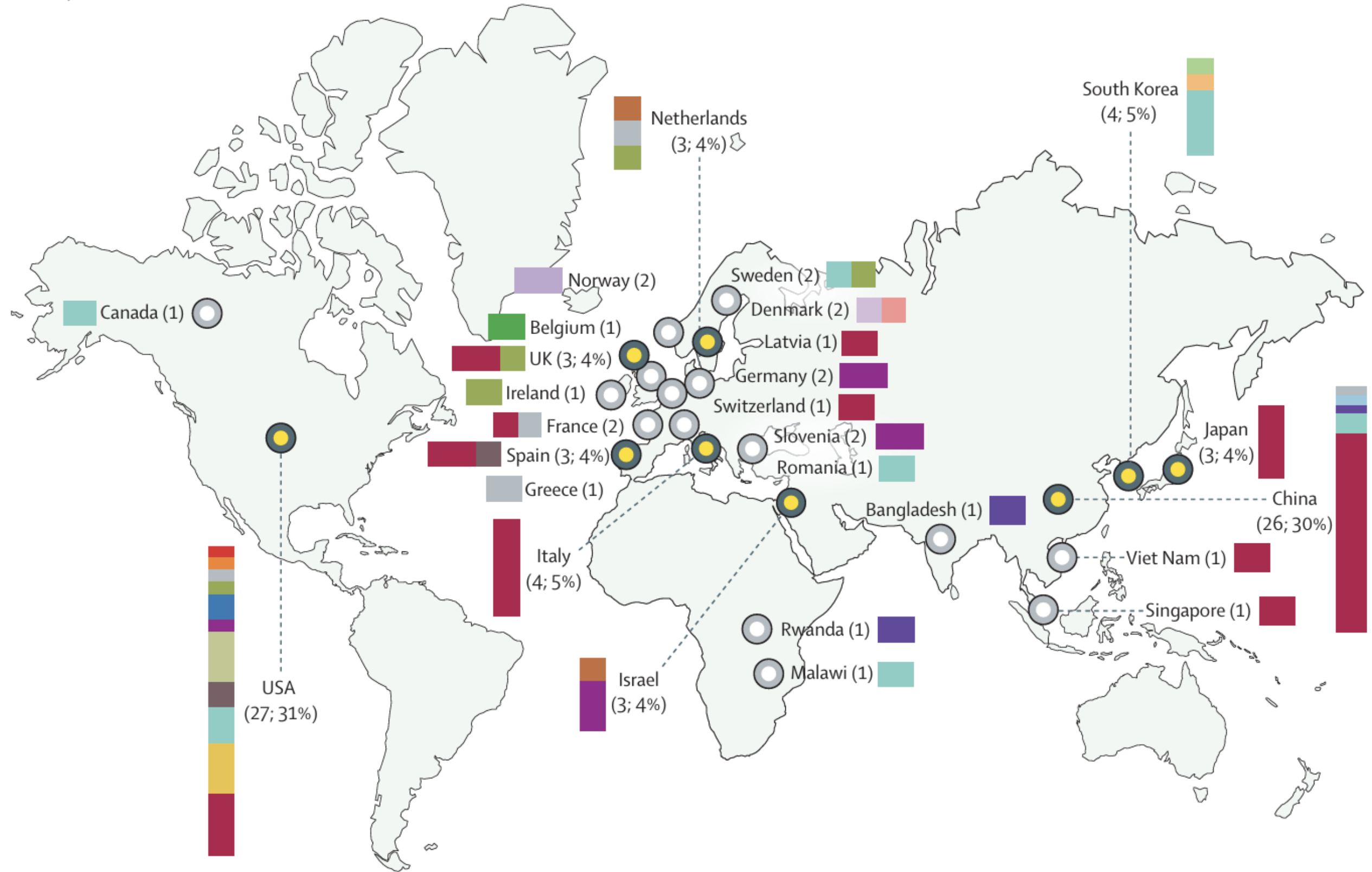




Overall distribution



Country-level distribution



# ¿Que tan bien funciona la IA técnicamente?



**86**

RCTs incluidos (2018–2023). El número más que se duplicó desde 2021.

**81%**

de los ensayos reportó resultado favorable para su endpoint primario

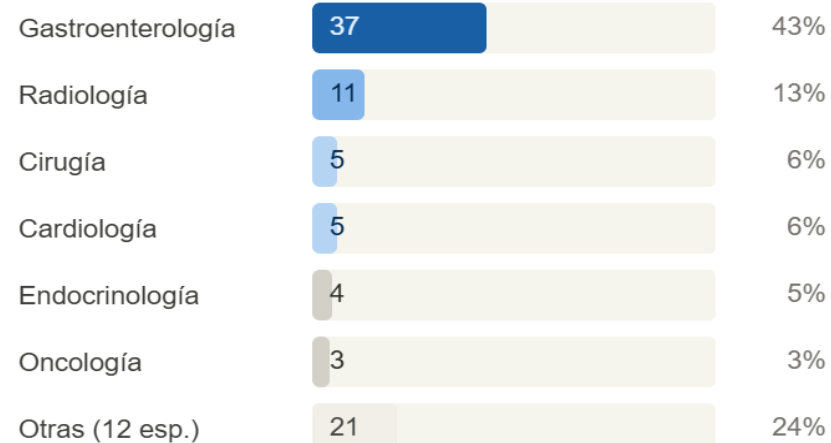
**359**

pacientes — mediana por ensayo (IQR 150–1050). Mayoritariamente estudios pequeños.

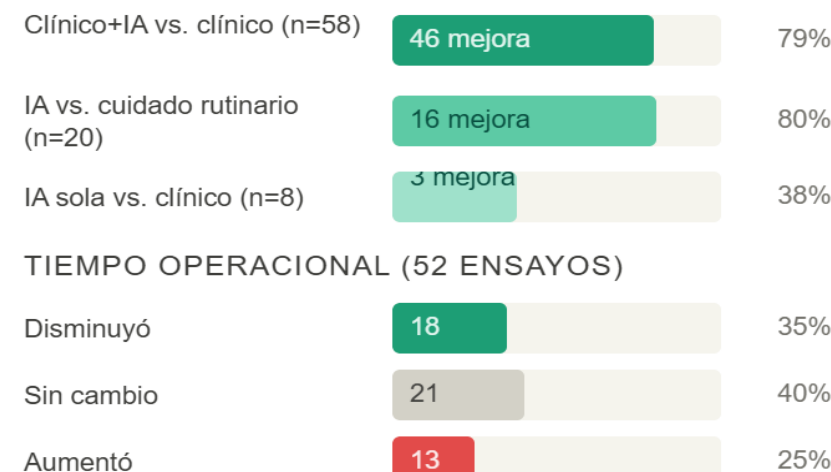
**63%**

fueron monocéntricos. Solo 7 ensayos multicéntricos, principalmente europeos.

## ESPECIALIDADES — DISTRIBUCIÓN DE LOS 86 RCTS



## RESULTADOS POR TIPO DE COMPARACIÓN



## PAÍSES CON MÁS ENSAYOS

**EEUU**  
27 ensayos  
31% · distribuidos en múltiples especialidades

**China**  
26 ensayos  
30% · 81% concentrados en gastroenterología

**Italia**  
4 ensayos  
5%

**Corea del Sur**  
4 ensayos  
5%

**España / UK / Japón / Israel**  
3 c/u  
4% cada uno

**América Latina**  
0 ensayos  
Ausencia total en la revisión

**Gastroenterología:** todos los ensayos evaluaron deep learning basado en video (colonoscopia/endoscopia). 65% provino de solo 4 grupos: U. Wuhan (8), Wision AI (6), Medtronic (6), Fujifilm (4). Radiología y oftalmología fueron las únicas especialidades que reportaron reducción consistente de tiempo operacional.

**Limitaciones críticas:** solo 19% de los ensayos desde 2021 citó las guías CONSORT-AI. Raza/etnia reportada en solo 26% de estudios, con 70,5% de participantes blancos en promedio — evidencia no extrapolable a población latinoamericana. Sesgo de publicación masivo: de 627 ensayos registrados en ClinicalTrials.gov, solo 9 (1%) estaban publicados.



# Gastroenterología — por qué es el área con más ensayos y resultados más variables

## Tipo de IA utilizada

- Exclusivamente **deep learning basado en video** (no imágenes estáticas) Redes neuronales entrenadas en millones de frames de colonoscopia.
- Operan en tiempo real durante el procedimiento: **analizan cada frame del video** y superponen un marcador visual cuando detectan una lesión potencial.
- Setup asistido: el sistema ayuda al clínico, no decide autónomamente.

## Para qué se usa

- **Detección de pólipos y adenomas** durante colonoscopia.
- Detección de lesiones tempranas en endoscopia alta (esófago, estómago).
- Clasificación automática de lesiones (neoplásica vs. no neoplásica).
- Mejora de calidad del examen para endoscopistas en formación.





# Gastroenterología — por qué es el área con más ensayos y resultados más variables

## Resultados favorables

- Tasa de detección de adenomas: **mejora significativa en la mayoría de ensayos** (ADR aumenta 4-6 puntos porcentuales en promedio).
- Tasa de pérdida de lesiones: reducción consistente.
- Tasa de éxito global: **78,4% de los 37 ensayos reportó mejora en endpoint primario.**
- Beneficio especial en endoscopistas con menos experiencia: el sistema actúa como "segundo par de ojos" que compensa la curva de aprendizaje.

## Resultados desfavorables

- Detección de adenomas avanzados clínicamente críticos: un metaanálisis de 21 ensayos mostró que la IA aumentó detección de pólipos pero **NO aumentó detección de adenomas avanzados de relevancia clínica.**
- Tiempo operacional: **38% de ensayos reportó aumento del tiempo de procedimiento.**
- Calidad heterogénea entre grupos: 65% de los ensayos provino de solo 4 grupos, lo que sugiere un campo dominado por desarrolladores con interés comercial directo (Medtronic, Fujifilm).
- ¿detectar más adenomas con IA reduce la incidencia de cáncer colorrectal a 5-10 años? Ningún ensayo tiene seguimiento suficiente para responderla aún.





# Radiología — por qué es el área con más ensayos y resultados más variables

## Tipo de IA utilizada

- Deep learning basado en imágenes estáticas (no video) CNNs para clasificación y detección de hallazgos en TAC, RX, resonancia y mamografía Modelos de segmentación para delimitar estructuras anatómicas o lesiones
- En algunos casos: sistemas de priorización de worklist (triaje de estudios urgentes)

## Para qué se usa

- Detección de nódulos pulmonares en TAC de tórax Análisis de radiografías de tórax (neumotórax, consolidaciones, cardiomegalia).
- Planificación de radioterapia (contorneo automático de órganos).
- Detección de lesiones en mamografía.
- Análisis de imágenes de fondo de ojo (retinopatía).
- Priorización de estudios urgentes en sala de urgencias.





# Radiología — por qué es el área con más ensayos y resultados más variables

## Resultados favorables

- Reducción del tiempo de informe: todos los ensayos de radiología que midieron tiempo operacional reportaron reducción significativa — es la única especialidad completamente consistente en este punto.
- Precisión diagnóstica: resultados favorables en la mayoría de aplicaciones específicas.
- Triage de urgencias: reducción del tiempo hasta diagnóstico en hallazgos críticos.

## Resultados desfavorables

- Solo 11 ensayos en el período — poca evidencia comparado con el volumen de aprobaciones regulatorias (más del 75% de los 1.451 dispositivos FDA son de radiología).
- Para radiología, otros diseños pueden ser más apropiados que RCTs: estudios pareados que comparan rendimiento del mismo médico con y sin IA en los mismos pacientes, eliminando confounders.
- La mayoría de la **evidencia publicada sobre radiología es retrospectiva** o usa endpoints de precisión, no outcomes clínicos del paciente.



# Outcome primarios

Wilhelm et al. · Lancet Reg Health Eur · Ene 2025

Contraste con Han et al. 2024

3.000

registros identificados → solo 19 estudios cumplieron criterios de inclusión

19

estudios incluidos: 18 RCTs + 1 cohorte prospectiva. Mediana 243 pacientes.

63%

reportó algún beneficio clínicamente relevante para el paciente

42%

documentó eventos adversos — los daños están sistemáticamente subregistrados



RIESGO DE SESGO — 18 RCTS EVALUADOS CON COCHRANE ROB 2

4

Bajo riesgo de sesgo

7

Algunas preocupaciones

7

Alto / serio riesgo de sesgo

## BENEFICIOS DOCUMENTADOS — ESTUDIOS CON BAJO RIESGO DE SESGO

### Significativo Mortalidad intrahospitalaria

Shimabukuro et al. — alerta temprana de sepsis (ML) en UCI  
9,0% vs 21,3% · p = 0,018  
Subgrupo sepsis: 13,6% vs 40,0% · p = 0,023

### Significativo Estancia hospitalaria

Shimabukuro et al. — mismo sistema de vigilancia  
Hospital: 10,3 vs 13,0 días · p = 0,042  
UCI: 6,3 vs 8,4 días · p = 0,030

### Significativo Síntomas depresivos

Delgadillo et al. — estratificación de cuidado en psiquiatría  
45,1% vs 52,3% mejora · p = 0,03  
Solo en casos no complejos (efecto no visto en casos severos)

### Significativo Tromboembolismo venoso

Huang et al. — sistema de apoyo a decisión clínica  
OR 0,55 · IC 95%: 0,34–0,88 · p = 0,01  
TVP: OR 0,50 · IC 95%: 0,30–0,83 · p = 0,02

### Significativo Readmisión hospitalaria

Wilson et al. — IA para derivación a cuidados paliativos  
60 días: OR 0,75 · IC 0,57–0,97 · p = 0,046  
90 días: OR 0,72 · IC 0,55–0,93 · p = 0,029  
No significativo a 30 días: OR 0,81 · IC 0,60–1,06

## OUTCOMES SIN BENEFICIO CONSISTENTE

### No confirmado Mortalidad — otros estudios

Bailey et al. — alerta deterioro clínico en sala médica general  
Sin diferencias significativas en mortalidad

### Mixto Mortalidad en oncología

Manz et al. — predicciones de mortalidad + nudge conductual  
Sin diferencias en muerte intrahospitalaria  
Sí redujo terapia sistémica al final de vida: OR 0,25 · IC 0,11–0,57 · p = 0,001

### No confirmado Estancia hospitalaria — otros

4 estudios adicionales no encontraron reducción significativa de estancia  
Beneficio solo replicado en 1 de 5 estudios que lo midieron

### No confirmado Ansiedad

Delgadillo et al. y Piette et al. — sin diferencias significativas en ansiedad  
p no significativa en ambos estudios

### No confirmado Exacerbaciones de asma

Seol et al. — sistema de IA para asma pediátrica  
12% vs 15% · p = 0,626 — no significativo

### Mixto Readmisión a 30 días

Wilson et al. — mismo estudio de paliativos  
OR 0,81 · IC 0,60–1,06 · p no significativa  
El beneficio aparece recién a los 60 y 90 días



# Outcome primarios



CONTRASTE DIRECTO — HAN ET AL. (2024) VS. WILHELM ET AL. (2025)

<b>86 RCTs</b> incluidos	vs	<b>19 estudios</b> incluidos
Mide <b>rendimiento técnico</b> (precisión, ADR, detección)	vs	Mide <b>outcomes del paciente</b> (mortalidad, morbilidad, calidad de vida)
<b>81%</b> de ensayos con resultado favorable	vs	<b>63%</b> con algún beneficio relevante — solo <b>2 de 5</b> sin sesgo alto
Predominio <b>gastro y radiología</b>	vs	Predominio <b>psiquiatría, oncología, medicina interna</b>
<b>América Latina: 0 ensayos</b>	vs	<b>América Latina: 0 ensayos</b>

**Hallazgo central de Wilhelm et al.:** de miles de estudios de rendimiento de IA en salud, solo 5 sin alto riesgo de sesgo midieron outcomes relevantes con algoritmo validado externamente — y solo 2 demostraron al menos un beneficio significativo: tratamiento de depresión (Delgadillo) y cuidados al final de vida (Manz). La evidencia existe pero es escasa, concentrada y poco generalizable.

**Brecha metodológica crítica:** solo 42% de los estudios documentó eventos adversos. La mayoría se enfoca en demostrar beneficios, ignorando la evaluación sistemática de daños. Para implementar IA con rigor clínico se requiere el mismo estándar que se exige a cualquier intervención médica: balance explícito beneficio-daño.

Wilhelm C, Steckelberg A, Rebitschek FG · Lancet Reg Health Eur. 2025;48:101145 · DOI: 10.1016/j.lanepe.2024.101145 · Open Access CC-BY · PROSPERO: CRD42023412156





# Gestión en Salud

Revisión sistemática · 20 estudios · 2019–2024

Revisión sistemática · 31 estudios · 2014–2024

85–95%

precisión de modelos IA para predicción de admisión hospitalaria vs. métodos tradicionales

0,81–0,93

AUC-ROC en los 7 estudios de mayor rigor metodológico para predicción desde triaje

48 h

ventana de predicción de hospitalizaciones con hasta 85% de precisión (modelos validados)

3 áreas

de impacto documentado: admisión, ocupación de camas y hacinamiento en urgencias

## APLICACIONES CON EVIDENCIA DOCUMENTADA

### Urgencias Predicción de admisión desde triaje

Modelos de ML predicen hospitalización en el momento del triaje, antes de la evaluación médica completa. Permite activar camas y recursos anticipadamente.

AUC 0,891 (IC 0,890–0,892) · Gradient Boosting Machine · mejor modelo reportado

### Camas Predicción de estadía hospitalaria (LOS)

Algoritmos que estiman días de hospitalización al ingreso, combinando datos de HCE, laboratorio y diagnóstico. Permiten planificación de alta y asignación de camas.

Random Forest y Neural Networks superan a modelos estadísticos clásicos en precisión

### Flujo Reducción de hacinamiento en urgencias

Sistemas que combinan predicción de admisión con NLP en registros no estructurados (notas de triaje, anamnesis) para optimizar flujo y asignación de recursos.

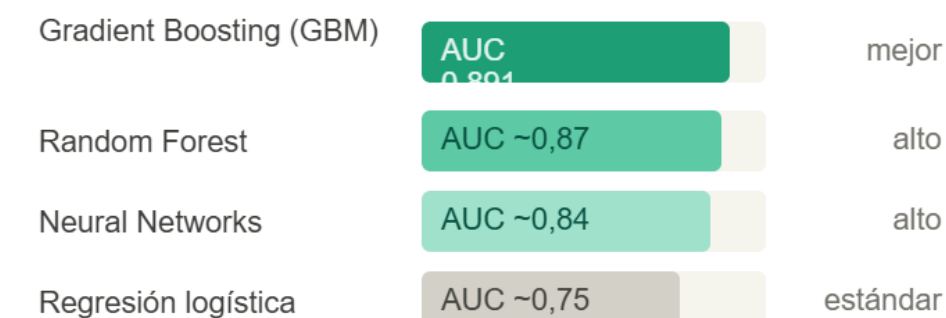
NLP en datos no estructurados mejoró significativamente el flujo de pacientes

### Triaje Triaje asistido por IA

Modelos de clasificación que apoyan la priorización de pacientes en urgencias. Superan a sistemas de triaje convencionales en predicción de outcomes.

Precisión de clasificación: 80,5% a 99,1% en estudios prospectivos (7 RS incluidos)

## RENDIMIENTO POR MODELO — PREDICCIÓN DE ADMISIÓN DESDE TRIAJE



## TIPO DE DATOS DE ENTRADA

### Datos estructurados de HCE

Signos vitales, laboratorio, diagnósticos previos, medicación, edad, sexo, comorbilidades. Base de la mayoría de los modelos.

### Datos no estructurados (NLP)

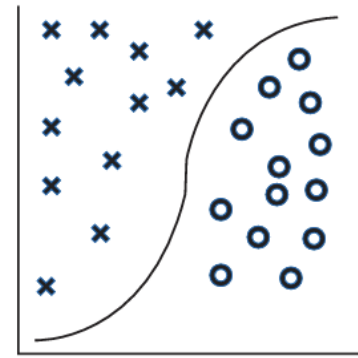
Notas de triaje, anamnesis libre, registros de enfermería. Los modelos que los incorporan muestran mejoras significativas adicionales en flujo y asignación.



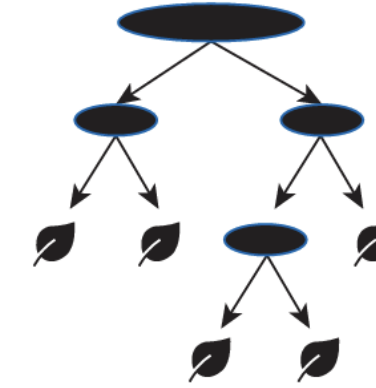


# Gestión en Salud

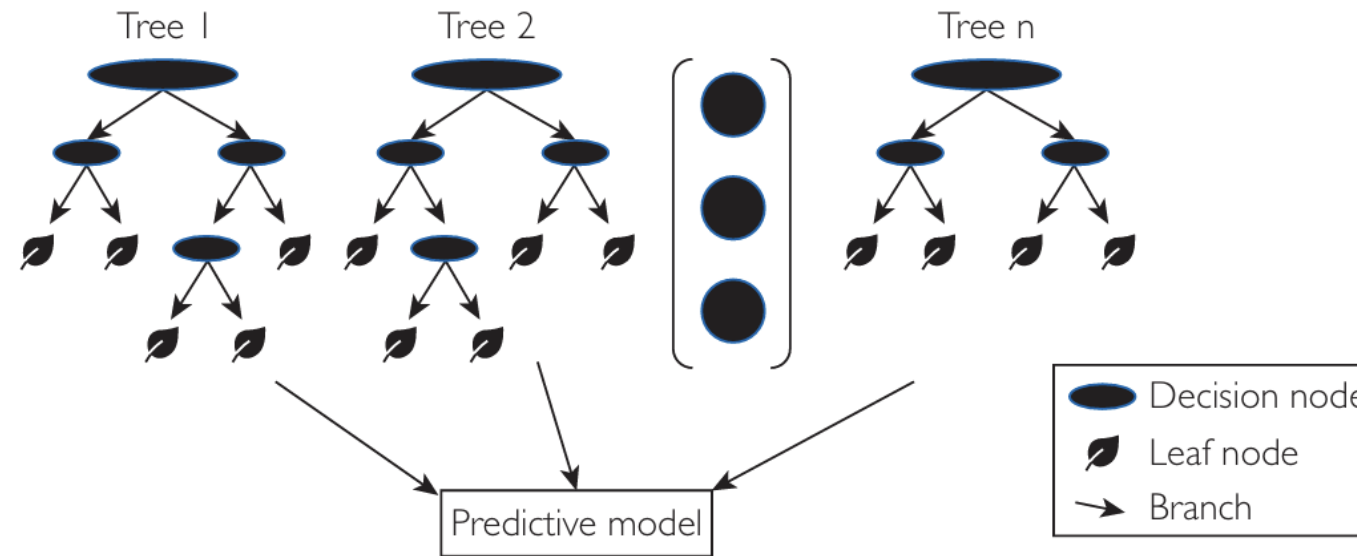
Logistic regression



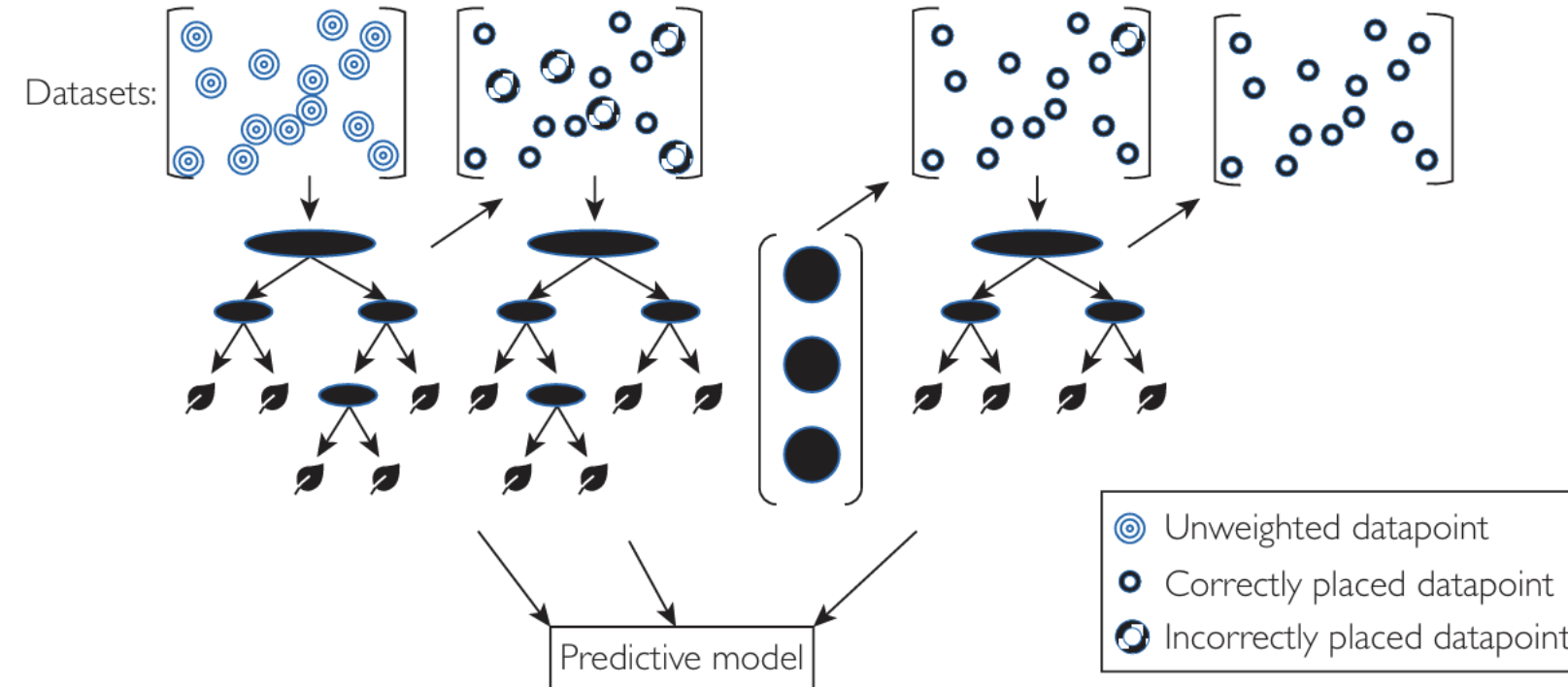
Decision tree



Random forest

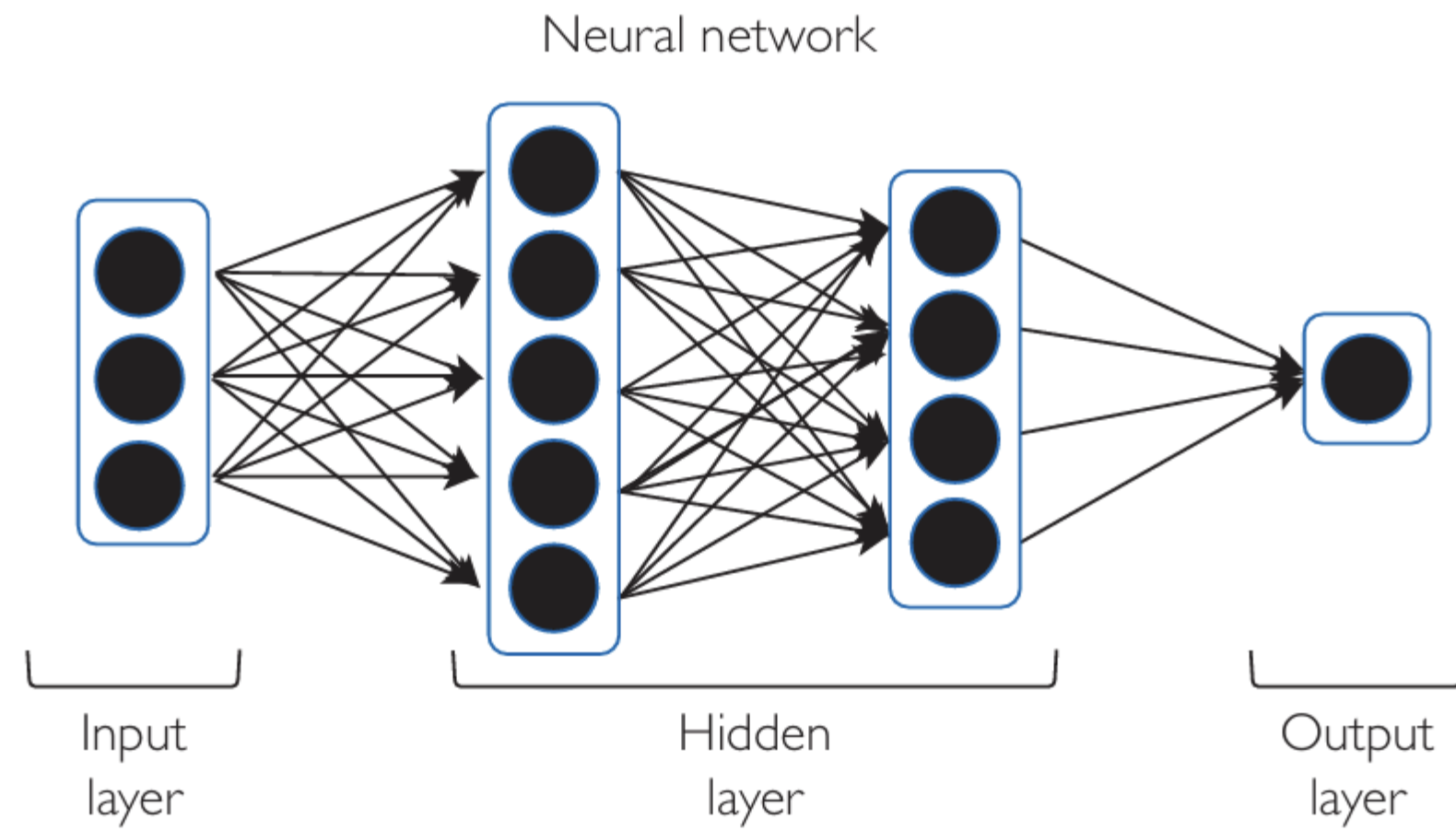


GBM and XGB





# Gestión en Salud





# Mensajes claves de la evidencia

## **La evidencia técnica es sólida, pero acotada**

La calidad técnica de los algoritmos está demostrada. Lo que no está demostrado es que eso se traduzca en mejores resultados para los pacientes.

## **Cuando se mide lo que realmente importa, la evidencia se adelgaza dramáticamente**

La brecha entre rendimiento técnico y beneficio clínico real no es un detalle metodológico — es el problema central.

## **La eficiencia operacional es prometedora, pero aún no tiene respaldo prospectivo**

Aún no sabemos con certeza si implementar estos sistemas en hospitales reales reduce efectivamente los tiempos de espera, optimiza las camas o mejora la experiencia del paciente en la práctica cotidiana.

## **Los daños están sistemáticamente ignorados**

En cualquier otra tecnología exigimos balance explícito beneficio-daño como condición para la implementación. La IA debería tener el mismo estándar.

## **El modo de uso determina el resultado, más que el algoritmo**

El factor humano (entrenamiento, integración, cultura organizacional) explica más varianza en los resultados que la calidad del modelo.

# Mensajes claves de la evidencia



La pregunta ya no es si la IA funciona, sino bajo qué condiciones

La síntesis de las tres revisiones lleva a una sola conclusión: la IA en medicina no es universalmente beneficiosa ni universalmente problemática. Es una tecnología cuyo impacto **depende críticamente de qué se mide, en qué población, con qué diseño, con qué entrenamiento del equipo y con qué integración al flujo de trabajo.**



# Paradigma vs herramienta





# Implicancias Éticas

## Responsabilidad legal

¿quién responde si la IA falla?  
médico · fabricante · hospital

86% de países cita incertidumbre  
legal como barrera principal  
OMS Europa 2025

EU AI Act: IA médica = alto riesgo  
supervisión humana obligatoria

el médico firma · el médico responde

## Sesgo y equidad

¿para quién funciona la IA?  
¿y para quién no?

0 ensayos latinoamericanos  
en 86 RCTs (Han et al. 2024)  
raza/etnia: solo 26% reportada

adoptar IA sin validación local  
= adoptar sus sesgos

el Colegio debe exigir validación local

¿qué implica  
la IA para  
los médicos?

## Formación médica

¿cómo se forma un médico  
que use la IA críticamente?

75% sin educación formal en IA  
en 192 facultades de medicina  
encuesta 4.500 estudiantes 2024

no programar · sí evaluar AUC,  
sesgos y límites del modelo

urgente: currículum y especialización

## Rol del médico

¿supervisor activo  
o firmante pasivo?

automation bias: confiar en  
el algoritmo sin cuestionar  
documentado en radiología y urgencias

3 modelos: IA-primero,  
médico-primero, asignación de casos

supervisar requiere comprender





Monitoreo fetal electrónico continuo (CTG) — evidencia acumulada vs. promesa original

**Premisa original (años 60): detectar sufrimiento fetal antes del daño neurológico**

problema demostrado: baja especificidad → muchas trazas "anormales" no predicen daño real → presión para intervenir sin beneficio

evidencia más sólida disponible — Cochrane 2017 (Alfirevic et al.) · 13 ensayos · 85.000+ mujeres

**Beneficio real**  
Convulsiones neonatales  
RR 0,50 (IC 0,31–0,80)

**Sin diferencia**  
Mortalidad perinatal  
Parálisis cerebral

**Daño documentado**  
Cesáreas: RR 1,63  
IC 95% 1,29–2,07

ensayo más grande disponible — Brocklehurst et al. · 46.042 mujeres · CTG computarizado vs estándar

**CTG computarizado (con algoritmo) vs CTG estándar: RR 1,01 (IC 0,82–1,25) para outcomes adversos**

Sin diferencia incluso eliminando la variabilidad inter-observador — el problema está en la señal, no en el intérprete

¿qué agrega la IA a este problema?

**Hereda el sesgo**  
Si se entrenó en datos  
donde traza anormal =  
cesárea, replica el patrón

**Más volumen de alertas**  
Monitoreo remoto las  
24h → más detecciones  
de baja especificidad

**Presión medicolegal**  
Alerta generada → no  
actuar = riesgo legal  
para el médico



# Ideas al cierre



## Lo que implica para la medicina

- El acceso a la IA no mejora automáticamente el razonamiento clínico. El estudio JAMA 2024 lo demostró con datos: médicos con GPT-4 disponible obtuvieron solo 2 puntos más que médicos sin IA, diferencia no significativa. **Lo que marca la diferencia es cómo se usa, no solo que esté disponible.**
- El médico sigue siendo responsable. El EU AI Act clasifica la IA médica como alto riesgo y exige supervisión humana. Usar IA sin entender sus limitaciones no transfiere responsabilidad — la multiplica.
- La resistencia bien fundada **no es lo mismo que el rechazo irracional.**
- **Los daños están sistemáticamente subregistrados.** Solo el 42% de los estudios de IA médica documenta eventos adversos. Antes de adoptar un sistema, debemos exigir datos de seguridad con el mismo estándar que exigimos a un fármaco.
- La IA no es el último eslabón de la cadena. El Epic Sepsis Model, Watson, el CTG continuo con IA remota: los tres fracasaron no porque el algoritmo fuera malo en sus datos de entrenamiento, **sino porque la integración al flujo clínico real no fue diseñada con rigor. La tecnología es el componente más fácil del problema.**



# Ideas al cierre



## Lo que deberíamos hacer

- Toda la evidencia viene de EEUU, China y Europa. **Ningún ensayo fue latinoamericano**. Implementar IA entrenada en esas poblaciones en Chile sin validación local es adoptar sus sesgos, sus supuestos y sus errores.
- El segundo argumento es impulsar formación en IA como competencia médica básica. El 75% de estudiantes de medicina no recibe educación formal en IA, mientras dos tercios de los médicos en ejercicio ya la usan.
- **Participar activamente en la gobernanza**, no solo como usuarios. La IA en salud concentra valor en un puñado de empresas globales. Chile pagará por tecnología que otros desarrollaron, sin participar en sus ganancias y sin poder auditar sus modelos.





**¡Muchas gracias!**



COLEGIO MÉDICO DE CHILE