

Papel y aplicación de las técnicas de inteligencia artificial en medicina

Alfredo Madrid García, PhD

Conflictos de interés



No tengo conflictos de interés que declarar

Agenda

- Motivación: ¿Dónde estamos?
- Introducción a la inteligencia artificial (IA) y al aprendizaje automático
- Ejemplo de aplicación de la IA en enfermedades musculoesqueléticas
- Otros: retos, ética, equidad, nuevas normativas
- Q&As



¿Quién soy?



- Doctor en Ingeniería Biomédica. Máster en Ciencia de Datos
- Investigador en el Servicio de Reumatología desde 2018
- Mi línea de investigación se centra en la aplicación de técnicas de procesamiento del lenguaje natural al texto libre de las notas clínicas.
- Descargo de responsabilidad: ***"I am not a physician"***

ORCID 0000-0002-1591-0467



[alfredo-madrid-garcia](https://www.linkedin.com/in/alfredo-madrid-garcia)



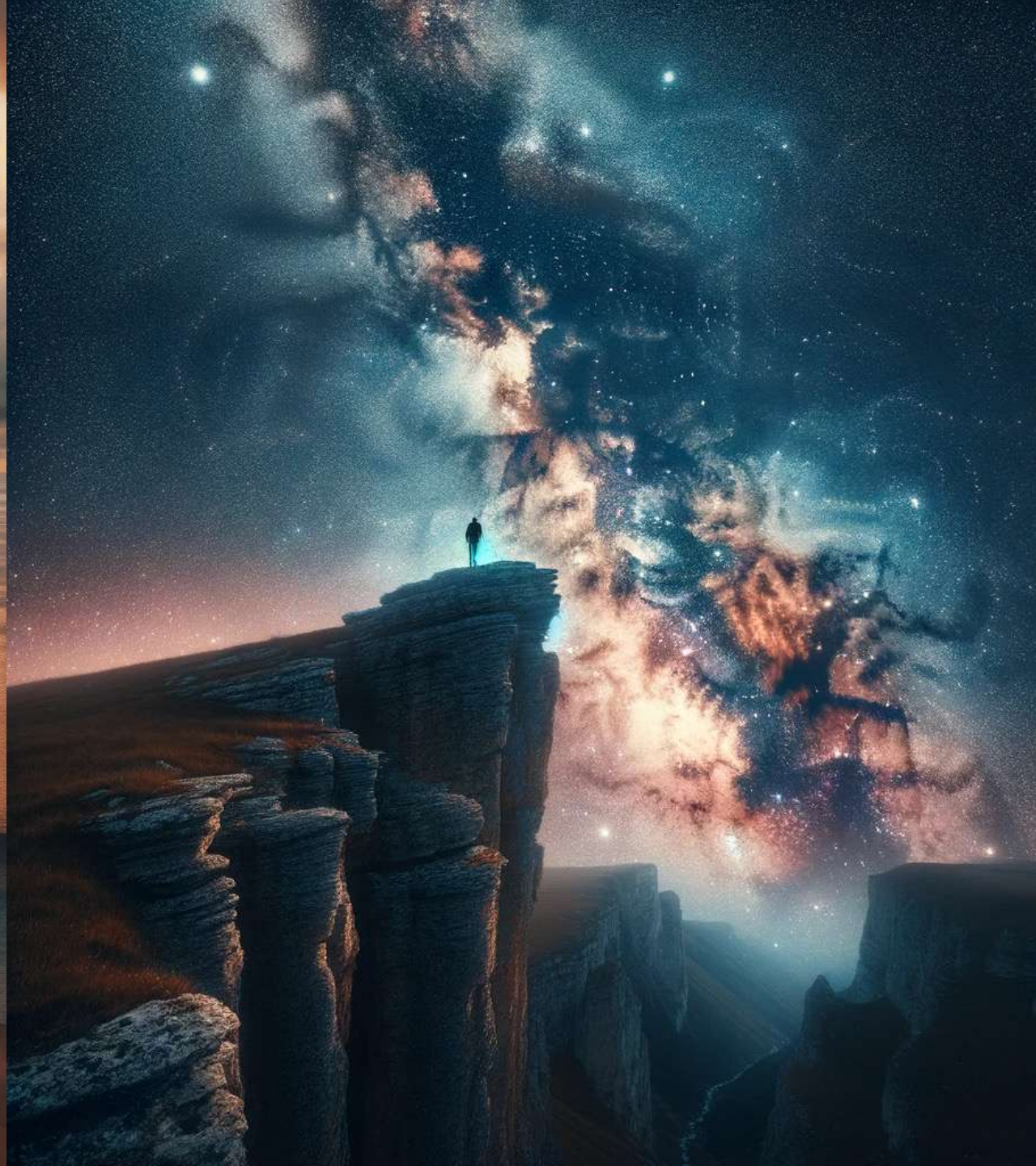
alfredo.madrid@salud.madrid.org



Motivación

De lo predictivo a lo generativo...





Los médicos de Madrid usarán un 'ChatGPT médico' para diagnosticar las enfermedades más raras. Empezarán por Atención Primaria

La Inteligencia Artificial cobrará mucha importancia en el diagnóstico de enfermedades raras

Sermas-GPT, Fundación 29

Tecnología ■

"CUESTIÓN DE VENDER MOTOS"

Malestar entre los médicos en Madrid por un proyecto de IA del que "nadie sabe nada"

El Servicio Madrileño de Salud (Sermas) anunció hace más de un mes un proyecto "pionero" para usar IA generativa en la detección de enfermedades raras. Los médicos y directores de centros de salud, sin embargo, no saben nada

Paloma, la inteligencia artificial que conversa con septuagenarios y detecta su soledad

La IA se queda lejos de ser número 1 MIR pero acierta el 73% del examen

Redacción Médica ha evaluado a esta nueva tecnología de la prueba para médico interno residente de 2024



La Comunidad de Madrid compra 223 nuevos ecógrafos para sus centros de salud que incluyen Inteligencia Artificial

Madrid apostará por la inteligencia artificial para diagnósticos médicos y pruebas de cáncer de mama

Esta medida sanitaria no busca reemplazar a los profesionales de la salud, sino complementar y mejorar la atención médica

SANIDAD DE MADRID

El Hospital Puerta de Hierro incorpora la inteligencia artificial para gestionar las urgencias

— Les ayudas a responder a detectar y responder los picos de asistencia

Ministerio y autonomías contienen la IA en sanidad con leyes y formación

Consejerías y Sanidad valoran sus claves para regular estas herramientas con control en el ámbito sanitario

SEIS: Sociedad Española de Informática de la Salud

Tabla 23: Proyectos tecnológicos implantados

Tendencias tecnológicas	Contestaron SI			En 2022		
	En 2019	En 2020	En 2021	Contestan SI	Contestan NO	No contestan
CCAA que tienen implantados proyectos de “Cloud Computing”	11	11	11	12	4	1
CCAA que tienen implantados proyectos de “Big Data”	10	9	11	10	6	1
CCAA que tienen implantados proyectos que integran datos de IOT en las HES (*)	–	–	–	3	13	1
CCAA que tienen implantados proyectos de Inteligencia Artificial (*)	–	–	–	8	8	1
CCAA que tienen implantados proyectos de Medicina de Precisión (*)	–	–	–	6	9	2
CCAA que tienen implantados proyectos que integran datos genómicos de los ciudadanos en las HES (*)	–	–	–	4	12	1

ÍNDICE SEIS 2022



Coordinadores:
Carlos García Codina
Francisco Martínez del Cerro
Alejandro Álvarez Cantalapiedra

MARZO 2023

The screenshot shows the homepage of The New England Journal of Medicine. At the top left is the journal's logo and name. To the right is a yellow 'SUBSCRIBE OR RENEW' button with an arrow and a small image of a document. Below the header is a navigation menu with links for SPECIALTIES, TOPICS, MULTIMEDIA, CURRENT ISSUE, LEARNING/CME, AUTHOR CENTER, PUBLICATIONS, and a SEARCH icon. The main content area features a large banner for 'AI in Medicine' with a collage of medical and technology images. Below the banner is a paragraph of text: 'Artificial Intelligence (AI) has tremendous potential to advance clinical practice and the delivery of patient care. A new Review article series, "AI in Medicine," explores the role of AI technology in clinical medicine and digital health, and examines the promise and pitfalls of its application across the health care continuum.' Below this text are three featured article cards. The first card on the left has an illustration of two people, one holding a smartphone and the other a tablet, with a blue arrow pointing to a laptop. The second card is titled 'HEALTH LAW, ETHICS, AND HUMAN RIGHTS' and 'Understanding Liability Risk from Health Care AI Tools' by M.M. Mello and N. Guha. The third card features a blue and red graphic and the 'NEJM AI' logo.

The NEW ENGLAND JOURNAL of MEDICINE

SUBSCRIBE OR RENEW →

SPECIALTIES TOPICS MULTIMEDIA CURRENT ISSUE LEARNING/CME AUTHOR CENTER PUBLICATIONS SEARCH

AI in Medicine

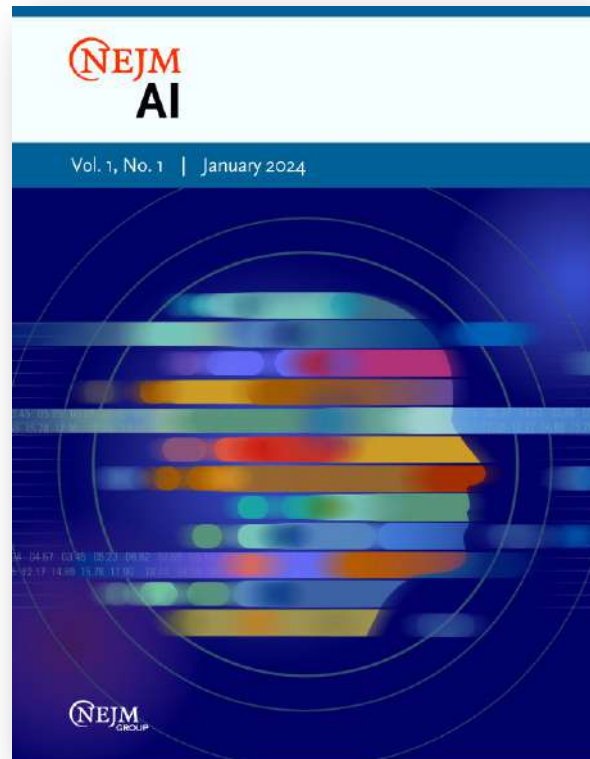
Artificial Intelligence (AI) has tremendous potential to advance clinical practice and the delivery of patient care. A new Review article series, "AI in Medicine," explores the role of AI technology in clinical medicine and digital health, and examines the promise and pitfalls of its application across the health care continuum.

HEALTH LAW, ETHICS, AND HUMAN RIGHTS
Understanding Liability Risk from Health Care AI Tools
M.M. Mello and N. Guha
The authors review challenges arising in malpractice litigation related to software errors to inform health care organizations and physicians about liability risk from AI adoption and about strategies to mitigate risk.

NEJM AI
NEJM Group, the publisher of the *New England Journal of Medicine*, has launched a new journal, *NEJM AI*, to identify and evaluate state-of-the-art applications of artificial intelligence to clinical medicine.
[Read free issue.](#)

Beam, Andrew L., et al. "Artificial intelligence in medicine." *New England Journal of Medicine* 388.13 (2023): 1220-1221.

COMING SOON NEJM AI – A NEW JOURNAL



29/2/24

*“NEJM AI is an interdisciplinary journal facilitating dialogue among stakeholders invested in using AI to transform medicine. NEJM AI intentionally pairs “pre-clinical” and clinical articles to deliver critical context to both clinicians and non-clinician researchers. The journal **bridges the fast-moving developments in AI, informatics, and technology in medicine with the application of these advancements to clinical practice.**”*

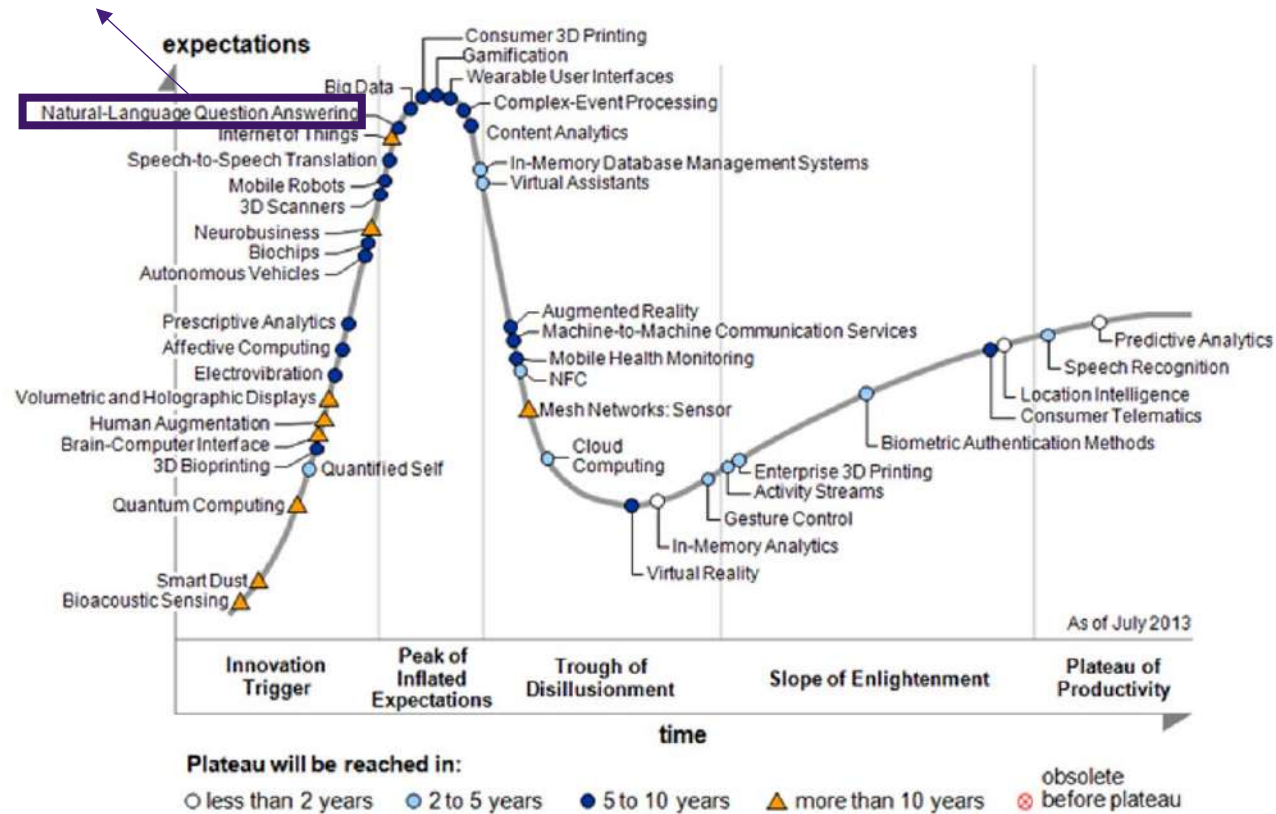
Volumen 1: January 2024

Ciclo de máxima expectativa de Gartner

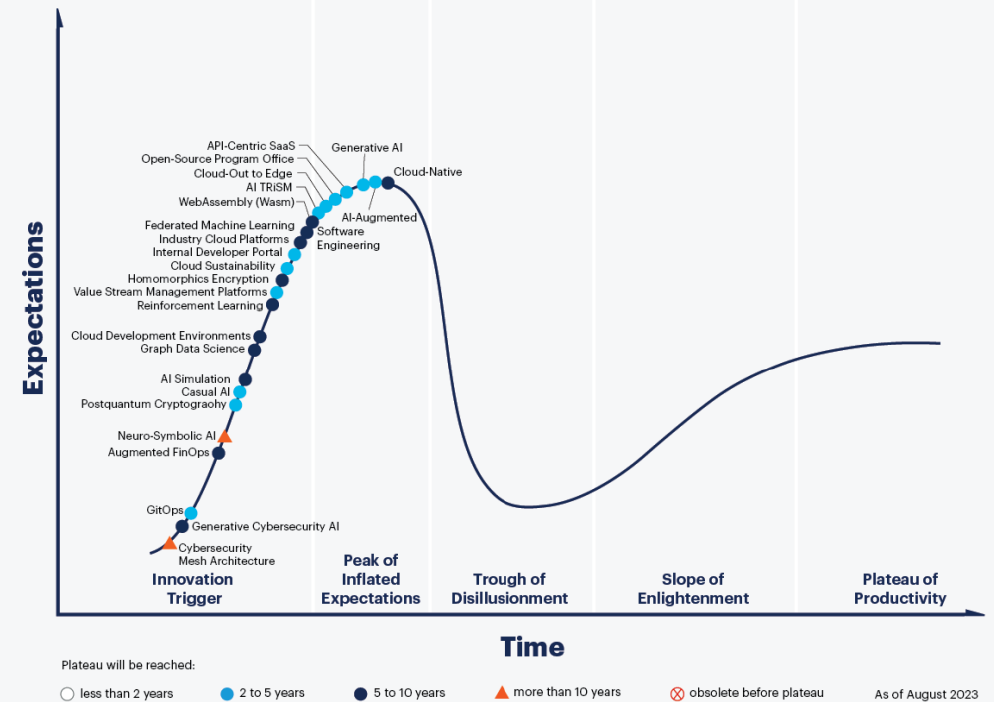
2013

2023

5-10 años

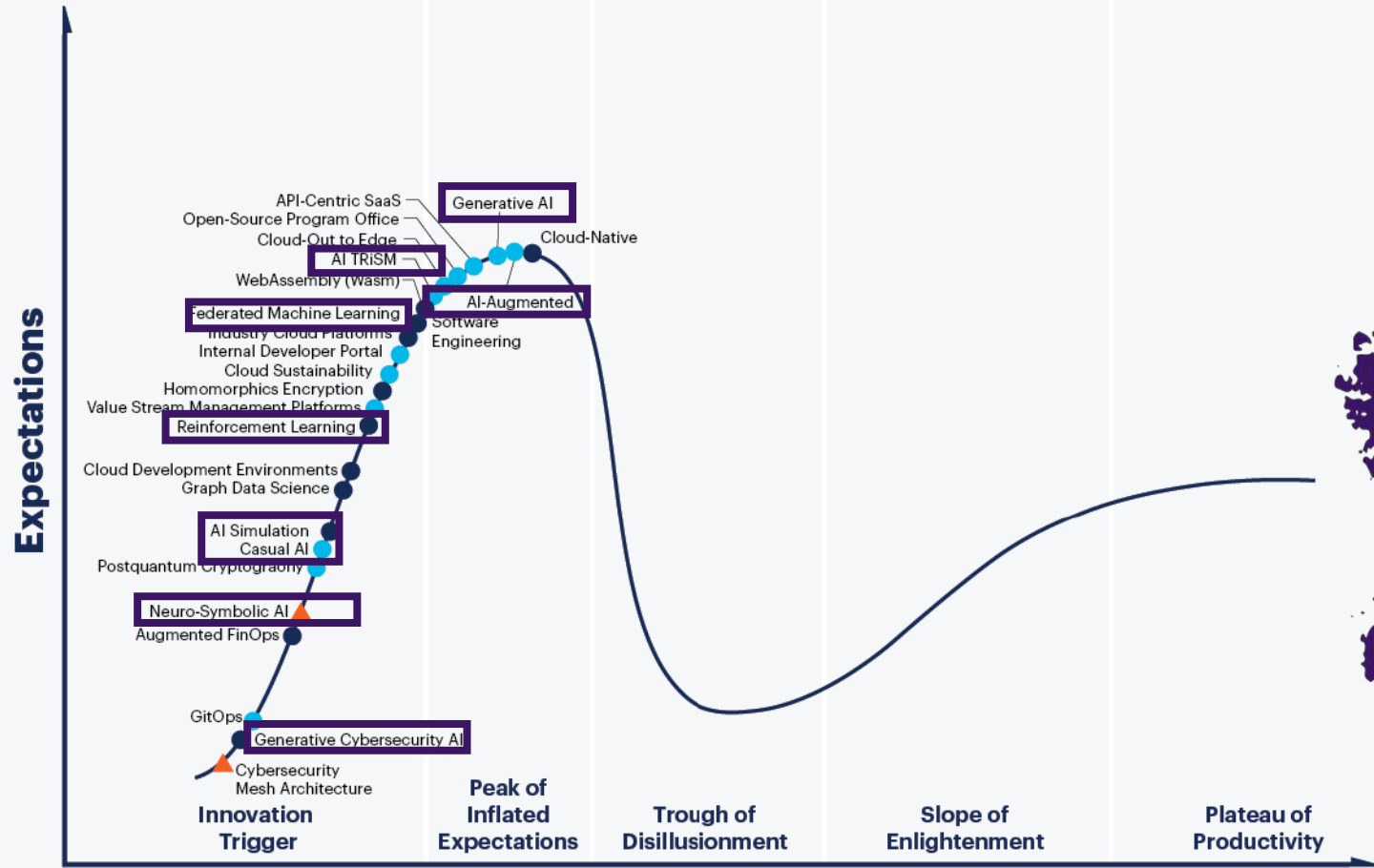


Hype Cycle for Emerging Technologies, 2023



¿Inteligencia artificial? ¿Machine learning?

Hype Cycle for Emerging Technologies, 2023



La IA generativa es una categoría de técnicas de IA en la que los algoritmos se entrenan en conjuntos de datos que pueden utilizarse para generar nuevos contenidos, como texto, imágenes o vídeo.

Plateau will be reached:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ⊗ obsolete before plateau As of August 2023

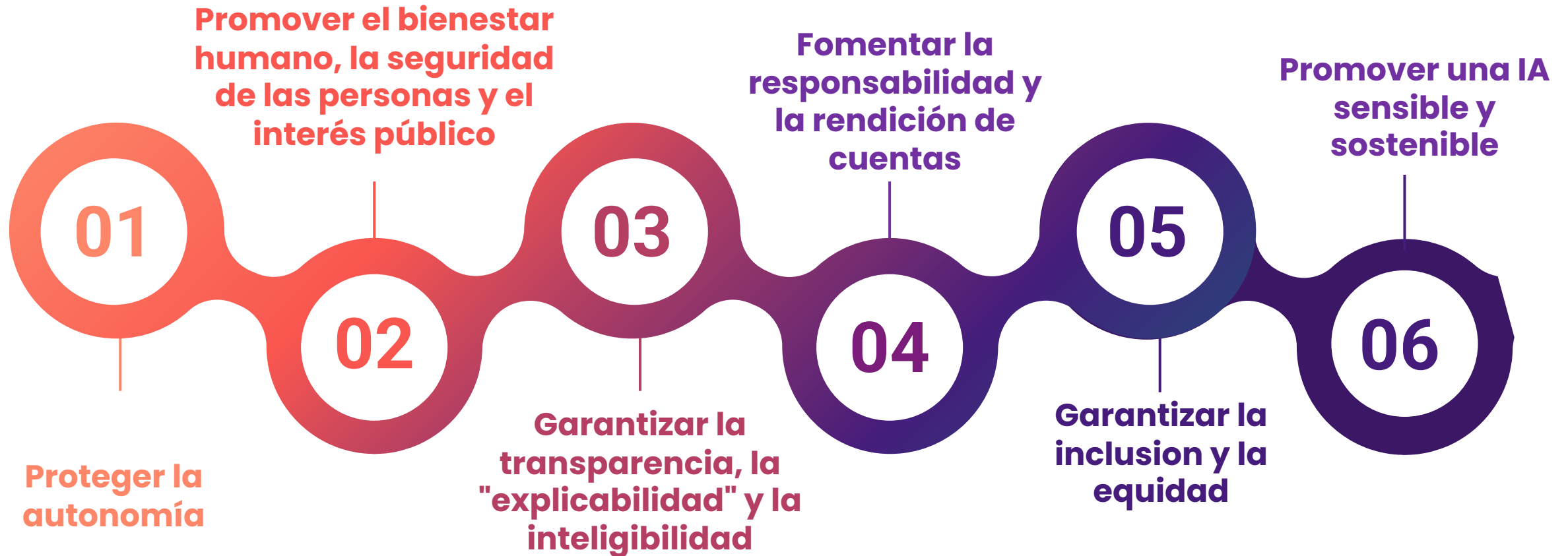
¿Dónde estamos?



AI can augment the ability of health-care providers to **improve patient care**, provide **accurate diagnoses**, **optimize treatment plans**, support pandemic preparedness and response, inform the decisions of **health policy-makers** or **allocate resources** within health systems

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance.

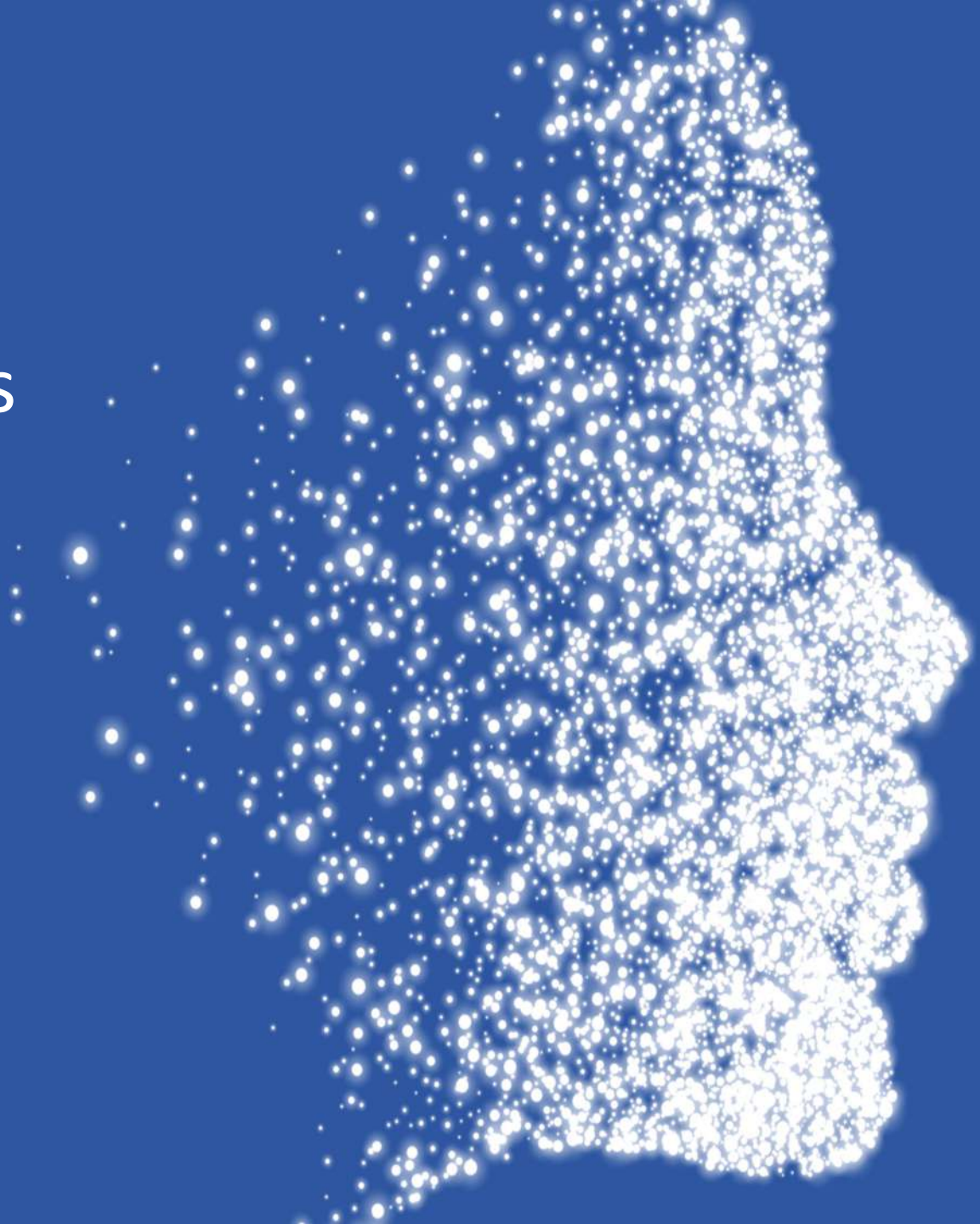
¿Dónde estamos?



¿Dónde estamos?

1. Los **humanos deben controlar** plenamente los sistemas sanitarios y las decisiones médicas
2. Las tecnologías de IA **no deben dañar a las personas** (ni mental ni físicamente)
3. Los **resultados** de la IA deben ser **comprensibles** para los desarrolladores, los usuarios y los reguladores
Compromiso entre la explicabilidad de un algoritmo (a costa de la precisión) y la **mejora** de la precisión
4. Exigencia de responsabilidad: El uso de tecnologías de IA en medicina **exige atribuir responsabilidades** dentro de sistemas complejos en los que la responsabilidad se distribuye entre numerosos agentes. **Cuando las decisiones médicas** de las tecnologías de IA **perjudican** a las personas, los procesos de responsabilidad y rendición de cuentas **deben identificar claramente el papel relativo de los fabricantes y los usuarios clínicos** en el daño. **Todos los agentes implicados en el desarrollo y despliegue de una tecnología de IA son responsables** *Everybody's problem becomes nobody's responsibility*
5. Uso y acceso **equitativos**, independientemente de la edad, el sexo, los ingresos, la capacidad u otras características.
 - Contratación de empleados de **diversas culturas** y disciplinas
 - Contemplar la diversidad de lenguas para que no haya **barreras de acceso**
 - **Sesgos raciales**
6. Si la tecnología de IA es ineficaz o genera insatisfacción **no debe usarse**

Di **HOLA** a los Grandes Modelos de Lenguaje



¿Dónde estamos?

Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models

, Camille Elepaño, Maria Madriaga,

Can you write me a report analyzing this chest X-ray?



La inteligencia artificial ChatGPT podría aprobar el examen para ejercer como médico en Estados Unidos

or near the passing
first to achieve this
milestone in AI
as able to achieve
m human trainers.

Findings:

- Devices: None.
- Lungs: No pneumothorax. No substantial pleural effusion. Lungs appear clear.
- Cardiomeastinal: Normal heart size. Mediastinal contours within normal limits.
- Other: No acute skeletal abnormality.

Impression:

No active disease seen in chest.

Enter a question here

Gemini



Perplexity

ANTHROPIC

Meta

Google Med-PaLM 2

<https://sites.research.google/med-palm/>

¿Dónde estamos?

BLOG ›

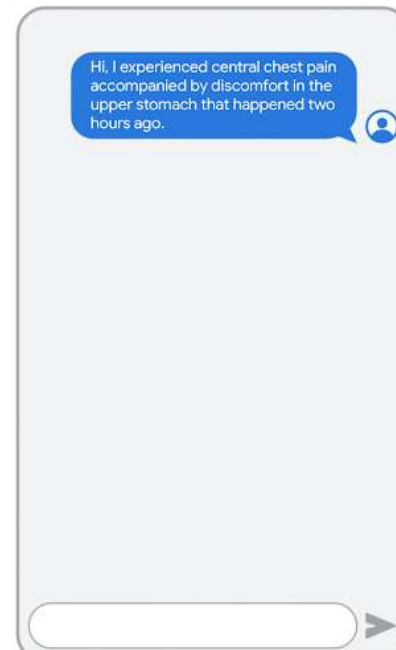
AMIE: A research AI system for diagnostic medical reasoning and conversations

Towards Conversational Diagnostic AI

Tao Tu^{*1}, Anil Palepu^{*1}, Mike Schaeckermann^{*1},
Khaled Saab¹, Jan Freyberg¹, Ryutaro Tanno², Amy Wang¹, Brenna Li¹, Mohamed Amin¹,
Nenad Tomasev², Shekoofeh Azizi², Karan Singhal¹, Yong Cheng², Le Hou¹, Albert Webson²,
Kavita Kulkarni¹, S. Sara Mahdavi², Christopher Semturs¹,
Juraj Gottweis¹, Joelle Barral², Katherine Chou¹, Greg S. Corrado¹, Yossi Matias¹,
Alan Karthikesalingam^{†1} and Vivek Natarajan^{†1}

¹Google Research, ²Google DeepMind

Conversation with AMIE



Sistema de IA optimizado para el diálogo diagnóstico basado en Med-Palm2

Fecha de publicación 2024-01-12

¿Dónde estamos?

Simulación de diálogo

Condition: Carpal Tunnel Syndrome

A 55-year-old female presents with a 3-month history of the following symptoms.

****Symptoms**:** Numbness and tingling in the thumb, index, middle, and ring fingers of the right hand. The numbness is worse at night and wakes her up from sleep. She also has pain in the hand and wrist that radiates up the forearm. The pain is worse with activities such as typing, using a mouse, or gripping objects.

****Past Medical History**:** Hypertension, hypothyroidism.

****Past Surgical History**:** None.

****Past Social History**:** Works as a data entry clerk.

****Patient Questions**:** What is causing my symptoms?

What can I do to make them better?

Doctor Agent (AMIE): So, how can I help you today?

Patient Agent (AMIE): Hi, I'm a 55-year-old woman and I've been having numbness and tingling in my right hand for the past 3 months. It's worse at night and it wakes me up from sleep. I also have pain in my hand and wrist that goes up my forearm. The pain is worse when I type, use a mouse, or grip objects. I'm wondering what's causing these symptoms and what I can do to make them better.

Doctor Agent (AMIE): I understand your concern. To better advise you, could you tell me:

- 1) Aside from the numbness and pain, have you noticed any weakness in your hand, like difficulty holding small items or clumsiness?
- 2) Any neck pain, or pain radiating down your arm beyond the forearm?
- 3) Do you have any medical conditions we should be aware of, and what medications do you take regularly?

Patient Agent (AMIE): 1. I have noticed some weakness in my hand. It's difficult to grip objects and I've dropped things a few times. 2. I don't have any neck pain or pain radiating down my arm beyond the forearm. 3. I have hypertension and hypothyroidism. I take levothyroxine and lisinopril for these conditions.

Doctor Agent (AMIE): What kind of work do you do? Does it involve repetitive hand movements, like typing or assembly? Have you had any injuries to your right hand or wrist in the past, even if long ago?

Patient Agent (AMIE): I work as a data entry clerk, so I do a lot of typing. I haven't had any injuries to my right hand or wrist.

Doctor Agent (AMIE): Aside from the numbness, pain, and weakness, have you noticed any swelling in your hand or wrist, especially in the morning?

Patient Agent (AMIE): I haven't noticed any swelling in my hand or wrist.

Doctor Agent (AMIE): This is helpful. Based on what you've shared, two possibilities come to mind. One is ****carpal tunnel syndrome**** which is common with repetitive hand use and causes the exact symptoms you describe. The other is ****cervical radiculopathy****, a nerve issue in the neck, but that usually has neck pain too.

¿Dónde estamos?

Patient Agent (AMIE): I see. What can I do to treat these conditions?

Doctor Agent (AMIE): For now, I'll focus on the more likely one, carpal tunnel. ****First****, wrist splinting, especially at night, is key. ****Second****, over-the-counter pain meds like ibuprofen can help. ****Third****, and this is crucial, ****ergonomic evaluation**** of your workstation is needed to reduce strain on your hand.

Patient Agent (AMIE): Thank you for the advice. I will try those things. I'll also look into getting an ergonomic evaluation of my workstation.

Doctor Agent (AMIE): Please do. ****If symptoms don't improve significantly in 4-6 weeks****, or if weakness worsens rapidly, an in-person evaluation is needed. A doctor can do nerve tests and confirm the diagnosis, and stronger options like steroid injections or surgery exist, but let's start conservative.

Patient Agent (AMIE): That sounds like a good plan. I'll follow your advice and see how things go. I'll be sure to contact a doctor if my symptoms don't improve or if they get worse. Thank you for your help.

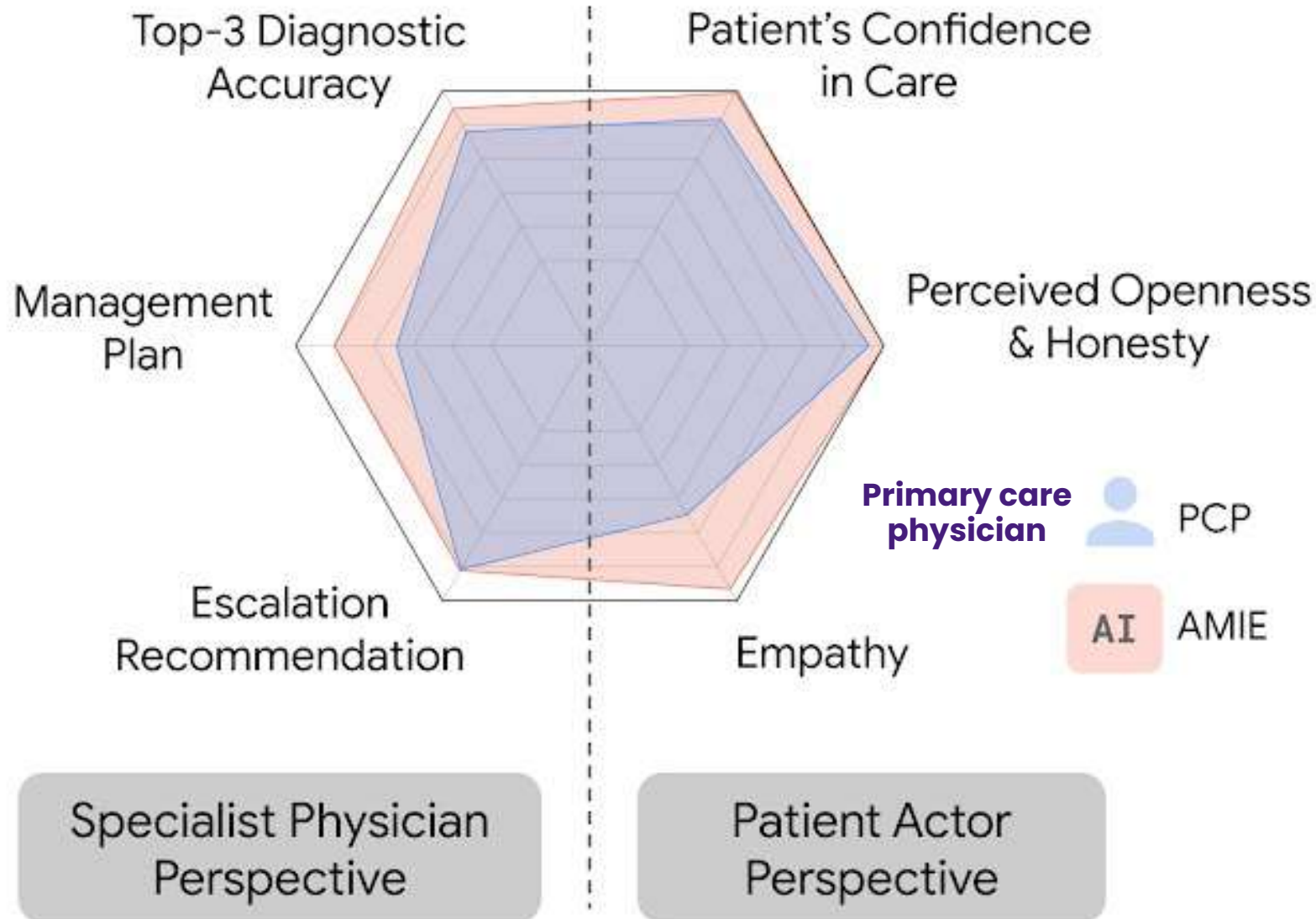
[Tu, T., Palepu, A., Schaekermann, M., Saab, K., Freyberg, J., Tanno, R., ... & Natarajan, V. \(2024\). Towards Conversational Diagnostic AI. *arXiv preprint arXiv:2401.05654*.](#)

Diagnóstico diferencial

- Síndrome del túnel carpiano
- Radiculopatía cervical

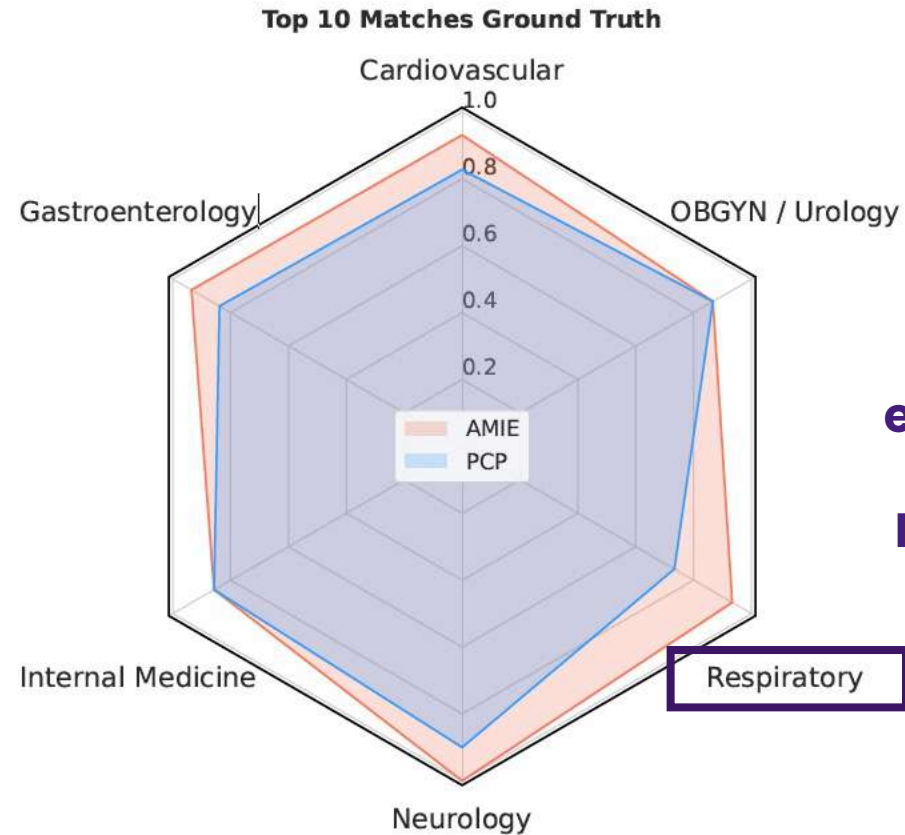
Plan de tratamiento

¿Dónde estamos?



- Estudio ciego en Canadá, UK e India
 - 149 casos
 - Comparación AMIE con respecto a 20 médicos de atención primaria
 - **AMIE es superior en 24/26 ítems.** De los otros dos ítems, se vio que **no era inferior**
 - **AMIE es superior en 28/32 ítems (perspectiva del médico)**

¿Dónde estamos?



Precisión a nivel de especialidad también superior

Neumología especialidad con mayor diferencia

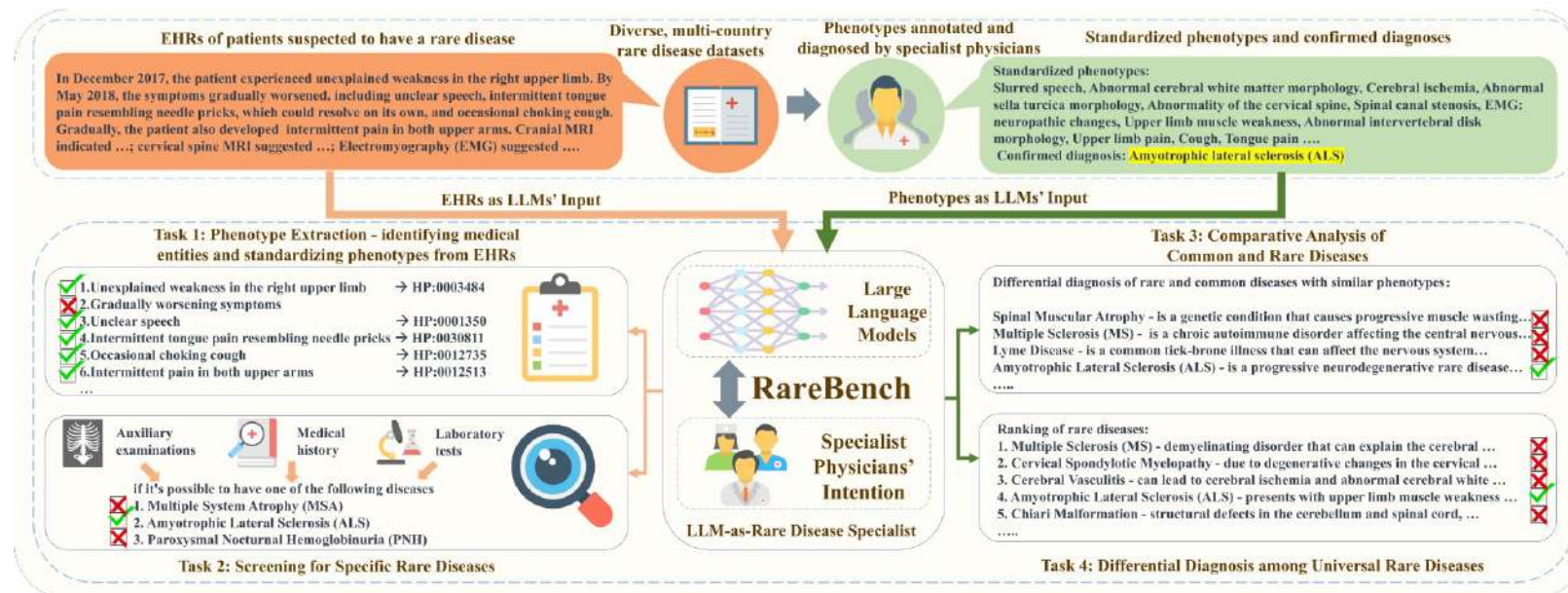
Figure A.8 | Specialist rated DDx accuracy by scenario specialty. Top 1/3/5/10 accuracy for scenarios of each specialty. Accuracies are based on the specialist ratings for AMIE and PCP differential diagnoses with respect to the ground truth. Number of dialogues per OSCE agent: Cardiology (29), Gastroenterology (31), Internal Medicine (14), Neurology (30), Respiratory (30), OBGYN / Urology (15).

¿Dónde estamos?

HEALTH

ChatGPT correctly diagnosed a 4-year-old's mysterious disease after 17 doctors failed

Kate Hull Sep 13, 2023, 12:20 AM CEST



Chen, X., Mao, X., Guo, Q., Wang, L., Zhang, S., & Chen, T. (2024). RareBench: Can LLMs Serve as Rare Diseases Specialists? *arXiv*. <https://arxiv.org/abs/2402.06341>

COMING SOON NEJM AI – A NEW JOURNAL











CURRENT ISSUE ▾ RECENTLY PUBLISHED PODCAST EVENTS AUTHOR CENTER ABOUT ▾ PUBLICATIONS ▾ 🔍

EDITORIAL



Why We Support and Encourage the Use of Large Language Models in *NEJM AI* Submissions

Authors: Daphne Koller, Ph.D. , Andrew Beam, Ph.D. , Arjun Manrai, Ph.D. , Euan Ashley, M.B., Ch.B., D.Phil. , Xiaoxuan Liu, M.B.Ch.B., Ph.D. , Judy Gichoya, M.B.Ch.B., M.S. , Chris Holmes, Ph.D. ,  for the editors and editorial board of *NEJM AI** [Author Info & Affiliations](#)

Published December 11, 2023 | *NEJM AI* 2023;1(1) | DOI: 10.1056/AIe2300128 | [VOL. 1 NO. 1](#)

*“We believe that **the use of LLM** tools can help scientists enhance the quality of their scientific work and **democratize both the creation and consumption of scientific knowledge**, thereby helping us maximally enable the scientific workforce to produce robust, novel scientific findings and disseminate them broadly”*

¿Dónde estamos?

“We predict that chatbots will be used by medical professionals, as well as by patients, with increasing frequency. Perhaps the most important point is that GPT-4 is not an end in and of itself. It is the **opening of a door to new possibilities as well as new risks. We speculate that GPT-4 will soon be **followed by even more powerful and capable AI systems**”**

SPECIAL REPORT AI IN MEDICINE

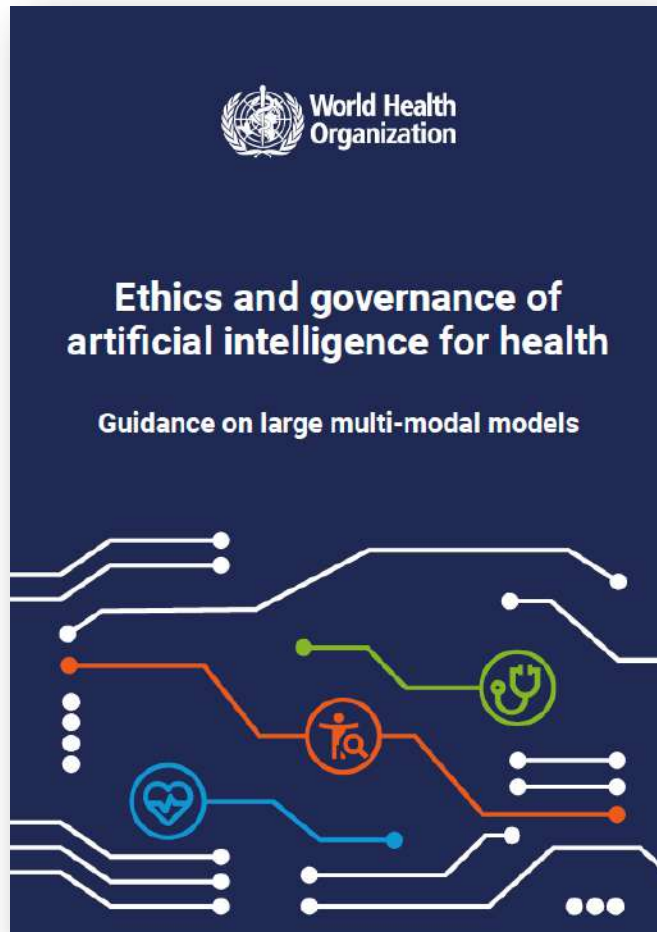
Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine

Peter Lee, Ph.D., Sebastien Bubeck, Ph.D., and Joseph Petro, M.S., M.Eng.

Article	Figures/Media	Metrics
11 References	284 Citing Articles	Letters
<p>THE USES OF ARTIFICIAL INTELLIGENCE (AI) IN MEDICINE HAVE BEEN GROWING IN MANY AREAS, including in the analysis of medical images,¹ the detection of drug interactions,² the identification of high-risk patients,³ and the coding of medical notes.⁴ Several such uses of AI are the topics of the “AI in Medicine” review article series that debuts in this issue of the <i>Journal</i>. Here we describe another type of AI, the medical AI chatbot.</p>		<p>March 30, 2023</p> <p>N Engl J Med 2023; 388:1233-1239 DOI: 10.1056/NEJMs2214184 Chinese Translation 中文翻译</p> <p>Editors</p> <p>Jeffrey M. Drazen, M.D., Editor, Isaac S. Kohane, M.D., Ph.D., Guest Editor, Tze-Yun Leong, Ph.D., Guest Editor</p>

Lee, Peter, Sebastien Bubeck, and Joseph Petro. "Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine." *New England Journal of Medicine* 388.13 (2023): 1233–1239.

¿Dónde estamos?



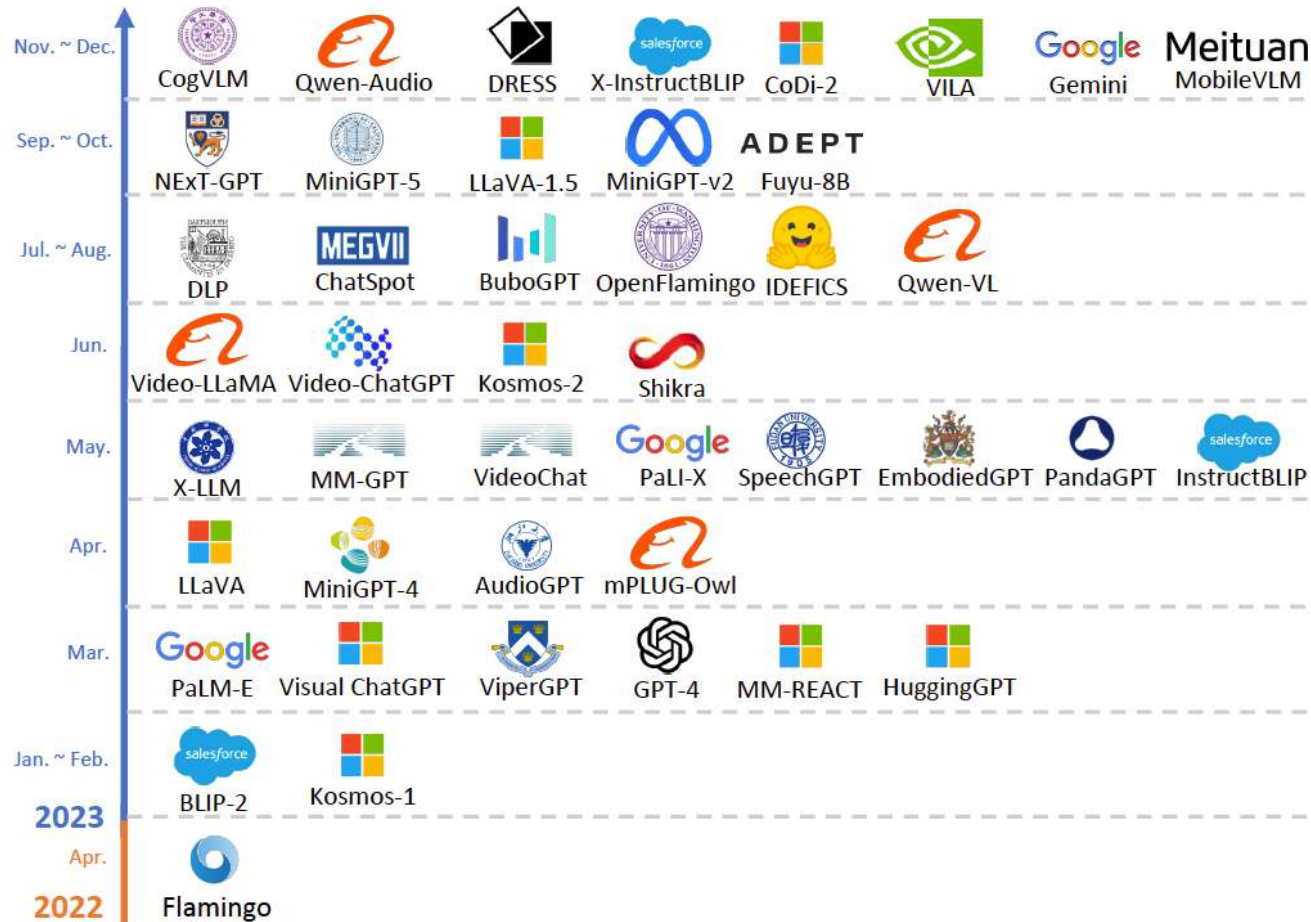
Fecha de publicación 2024-01-18

Multimodal: pueden aceptar más de un tipo de dato entrada (e.g., Imagen + Texto)

*It has been predicted that LMMs will have **wide use and application in health care**, scientific research, public health and drug development. LMMs are also known as “general-purpose foundation models”*

World Health Organization. (2024). Ethics and governance of artificial intelligence for health: guidance on large multi-modal models.

¿Dónde estamos?



Zhang, D., Yu, Y., Li, C., Dong, J., Su, D., Chu, C., & Yu, D. (2024). MM-LLMs: Recent Advances in MultiModal Large Language Models. arXiv preprint arXiv:2401.13601.

Figure 1: The timeline of MM-LLMs.

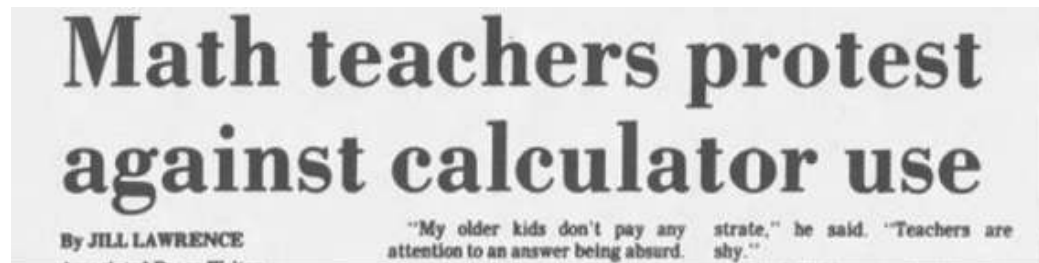
¿Dónde estamos?

Beneficios potenciales del uso de LLMs en salud

- 1 Ayuda diagnóstico y atención clínica
- 2 Uso orientado al paciente
- 3 Tareas administrativas
- 4 Educación personal clínico
- 5 Investigación y desarrollo de fármacos

Riesgos potenciales del uso de LLMs en salud

- 1 Sobreestimación de los beneficios
- 2 Accesibilidad y asequibilidad
- 3 Sesgos
- 4 Impacto en el trabajo
- 5 **Dependencia del Sistema sanitario**



¿Dónde estamos?

Los gobiernos deben apoyar el desarrollo colectivo de **normas internacionales** para la gobernanza de los LMM y otras formas de IA utilizadas en la **asistencia sanitaria**

- 1 Crear una agencia pública de investigación internacional con financiación y recursos humanos para llevar a cabo grandes proyectos transformadores cuyos **resultados se compartan abiertamente**. Dicha entidad podría encargarse de desarrollar las formas más avanzadas y **arriesgadas** de inteligencia artificial en **instalaciones de alta seguridad, lo que ilegalizaría otros intentos de crearlas**.
- 2 Líderes mundiales y ejecutivos del sector tecnológico han pedido que la **IA reciba un trato similar** al de las **armas nucleares**, con un marco regulador mundial similar al de los tratados sobre el uso de armas nucleares.

¿Dónde estamos?

First AI-generated drug enters human clinical trials, targeting chronic lung disease patients

By Melissa Rudy, Fox News

Published June 29, 2023, 1:17 a.m. ET

<https://www.clinicaltrials.gov/study/NCT05938920>

The discovery, a world first, took less than 18 months **and cost 10% as much as a conventional program.** <https://www.nature.com/articles/d43747-021-00039-5>

¿Dónde estamos?

- **Fibrosis pulmonar idiopática**
 - **Fase II**
 - **Intervención**
 - **Aleatorizado**
 - **Doble ciego**
- **Frente a placebo**

Study Overview

Brief Summary

The goal of this clinical trial is to learn about INS018_055 in adults with Idiopathic Pulmonary Fibrosis (IPF).

The primary objective is to evaluate the safety and tolerability of INS018_055 orally administered for up to 12 weeks in adult subjects with IPF compared to placebo.

Official Title

A Phase IIa, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety, Tolerability, Pharmacokinetics, and Efficacy of INS018_055 Administered Orally to Subjects With Idiopathic Pulmonary Fibrosis (IPF)

Conditions

Idiopathic Pulmonary Fibrosis (IPF)

Intervention / Treatment

- Drug: INS018_055
- Drug: Placebo

Other Study ID Numbers

- INS018-055-003

Study Start (Actual)

2023-06-19

Primary Completion (Estimated)

2024-05-25

Study Completion (Estimated)

2024-06-11

Enrollment (Estimated)

60

Study Type

Interventional

Phase

Phase 2

¿Dónde estamos?

AI 100: The most promising artificial intelligence startups of 2023. CBINSIGHTS



AI development tools

AI chips & processors Saliense Labs, XANADU	Synthetic data hazy, MOSTLY.AI	NLP annotation Argilla	Computer vision visualization rerun.io	Federated learning platforms APHERIS, DynamofL	Version control & experiment tracking Weights & Biases
Vector database tech LangChain, Weaviate, Zilliz	ML development & deployment Lightning, mindsdb, mosaic, feplicate	Model validation & monitoring arize, LatticeFlow, WHYLABS	AI auditing & governance credo.ai, Holistic AI	Foundational models & APIs AI21labs, cohere, contextual.ai, Hugging Face, OpenAI	

Cross-industry applications

Synthetic voice ElevenLabs	Image & text generation Jasper, Midjourney	Privacy & security HIDDENLAYER, Opaque, Protect AI, RELYANCE AI	Code generation diffblue, replit, warp	Sales & customer support AssemblyAI, PolyAI, talka, tavus	AI assistants & HMIs A DEPT, ANTHROPIC, Inflection, Wispr
Emotion analytics hume	Productivity tools mem	Warehouse & logistics covariant	Content moderation Logically, unitary	Smell tech osmo	Translation DeepL
General-purpose humanoids FIGURE, SANCTUARY AI	Image processing Metaspectral, VISIONARY.AI	Digital twins blackshark.ai	Design tools Galileo AI, Poly	Climate tech muir.ai	Quantum AI software Sandbox AI
Search Perplexity, Twelve Labs, vectara, YOU					

Industry-specific

Materials & manufacturing Cradle, MakinaRocks	Gaming convai, MODULATE	Fashion & retail refiberd, see1	Energy phaidra, WeaveGrid	Healthcare QEMIA, Atomic AI, ATROPOSHEALTH, Generate:Diomedices, navina, SURTLER MEDICAL, turbine, XtalPI
Defense VANNEVAR Labs	Finance arteria, Boosted.ai	Agriculture BEEWISE, VERDANT ROBOTICS	Physical infrastructure exadigo	
Education Ello	Media & entertainment character.ai, descript, runway, wonder	Legal Harvey	Auto & mobility helm.ai, PHANTOM AI	Construction Augmenta

Note: Companies are private as of 6/20/23.

¿Dónde estamos?



4/9 empresas más prometedoras están orientadas a la elaboración de nuevos fármacos

AQUEMIA

ATOMIC AI

XtalPi

Generate:biomedicines

Desarrollo de fármacos

- **Aquemia:** descubrir candidatos a **fármacos innovadores** más rápidamente
- **Atomic AI:** liberar el poder de la inteligencia artificial en la estructura molecular para construir fármacos basados en **ARN**, creando medicamentos que puedan abordar **enfermedades que antes eran intratables**.
- **XtalPi:** descubrimiento de fármacos
- **Generate:Biomedicines:** generar medicamentos bajo demanda en múltiples modalidades terapéuticas con una velocidad y tasas de éxito sin precedentes, lo que representa un cambio fundamental en el descubrimiento de fármacos.

¿Dónde estamos?

Atropos health

Navina

Subtle medical

Turbine

Otros

- **Atropos Health:** transforma datos médicos en evidencia del mundo real, **estructura datos médicos**
- **Navina:** reduce la **carga de revisión de las historias clínicas** de los médicos en un 30 % y el *burn out*
- **Subtle Medical:** mejorar la **calidad de la imagen** médica de **MRI y PET**
- **Turbine:** simulación con aprendizaje automático para mapear y **modelar cómo miles de proteínas de señalización** interactúan caracterizando el comportamiento **a nivel celular del cáncer y la respuesta al tratamiento**




¿Dónde estamos?

AQEMIA Announces a Major Multi-year Collaboration of \$140 Million With Sanofi

This collaboration aims at accelerating the discovery of small molecules across various therapeutic areas, leveraging its unique Deep Physics platform combined with Generative AI. AQEMIA to receive up to a total of \$140M in upfront and milestones across all programs.

Expanding Amgen Collaboration:

Generate:Biomedicines will highlight the expansion of its existing collaboration agreement with Amgen (NASDAQ: AMGN). Amgen has exercised its rights under the collaboration agreement to opt in for a sixth program. This represents the first expansion of the original agreement, and as a result, Amgen will make an undisclosed upfront payment and will pay up to \$370 million in future milestones and royalties up to low double digits for this new program.



Turbine Announces Collaboration to Uncover Biological Mechanisms of Drug Resistance in Hematological Cancers

Turbine, a company that uses cell simulations to guide biopharma R&D and bring the right drug to every patient in need, has successfully added €5.5 million from MassMutual Ventures (MMV) to its recently closed Series A financing round. This investment brings the company's oversubscribed Series A round to €25.5 million and adds further U.S.-based investors to the company's financing syndicate. Alongside the funding, Turbine announced the appointments of four independent directors to its board.

Turbine Establishes Research Collaboration with Ono Pharmaceutical to Identify and Validate Novel Oncology Targets

Turbine, a company that uses cell simulations to guide biopharma R&D and bring the right drug to every patient in need, has successfully added €5.5 million from MassMutual Ventures (MMV) to its recently closed Series A financing round. This investment brings the company's oversubscribed Series A round to €25.5 million and adds further U.S.-based investors to the company's financing syndicate. Alongside the funding, Turbine announced the appointments of four independent directors to its board.

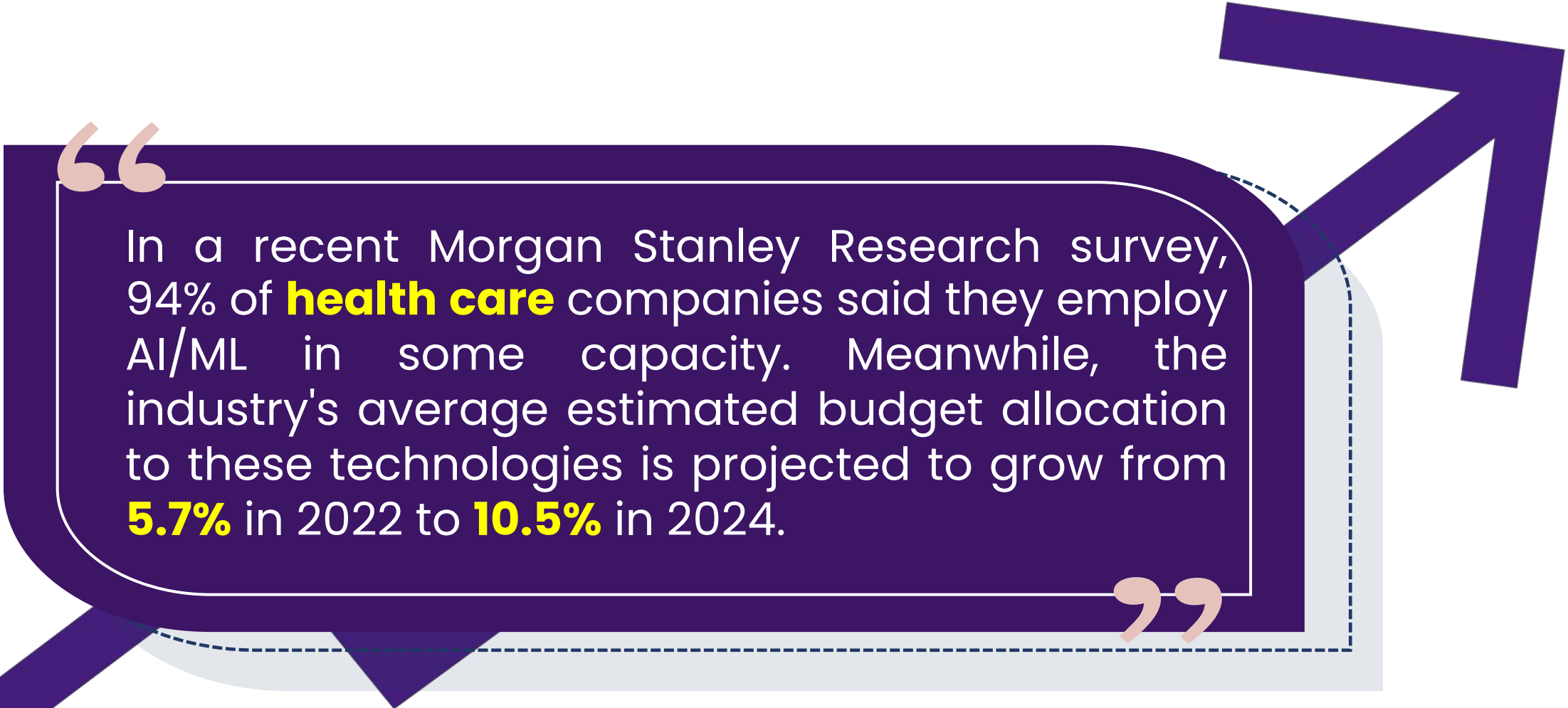
XtalPi in the News



2023-06-01
XtalPi Announces Collaboration with Lilly, Using AI + Robotics to Uncover First-in-class Therapeutics
[Read More >>](#)

2023-01-11
XtalPi Names New COO for U.S. Operations, Bringing Its Autonomous Labs to Boston in Continued Global...
[Read More >>](#)

2022-10-27
XtalPi Announces Research Collaboration with Janssen
[Read More >>](#)



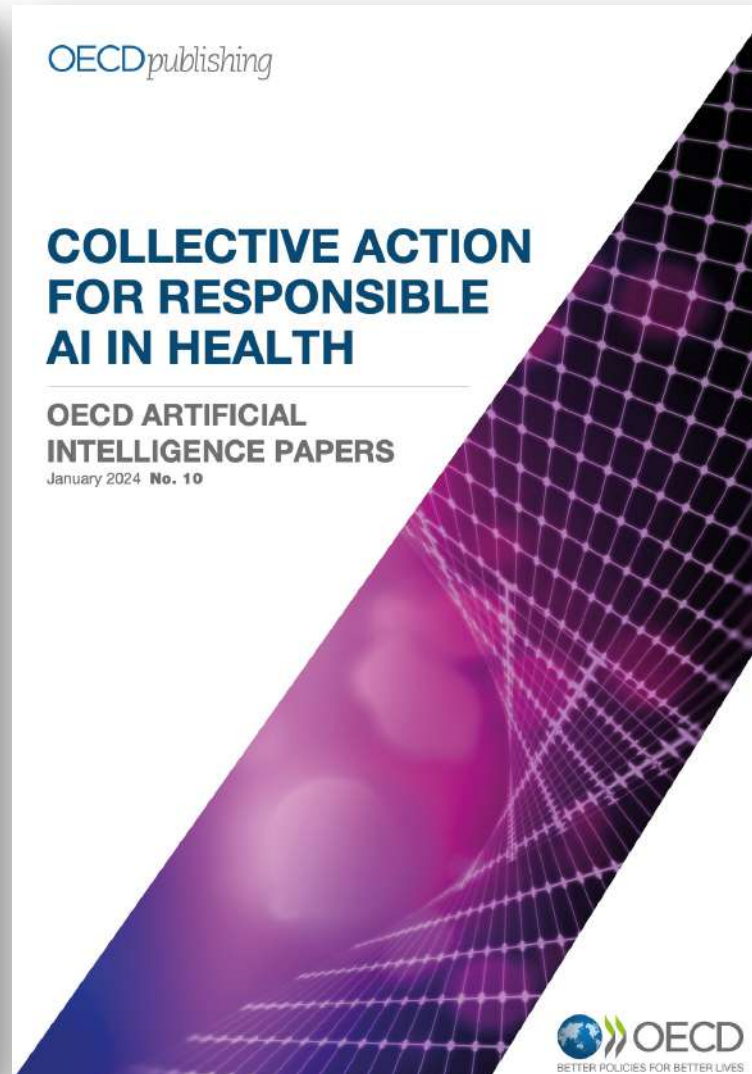
In a recent Morgan Stanley Research survey, 94% of **health care** companies said they employ AI/ML in some capacity. Meanwhile, the industry's average estimated budget allocation to these technologies is projected to grow from **5.7%** in 2022 to **10.5%** in 2024.

[How Artificial Intelligence Could Reshape Health Care](#)

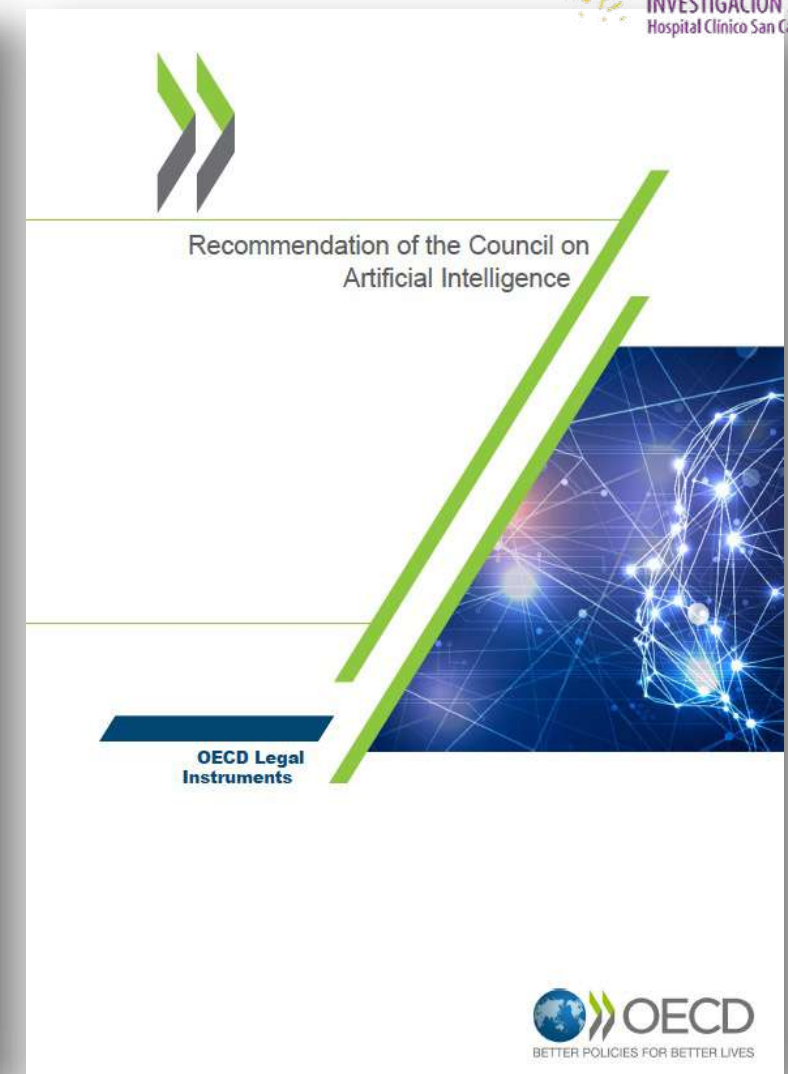


Fecha de publicación 2024-01

29/2/24



Fecha de publicación 2024-01-19



Fecha de publicación: 2019

**AI *saves lives, helps* the health
workforce, and *unlocks value***

**AI can make health
systems more
people-centred**

**AI can help health
professionals
provide more time
to care**

**AI solutions protect
personal data while
improving health
outcomes**

**AI can help to protect
digital health
infrastructure from
security threats**

There are risks that must be effectively addressed...

...likewise, there are risks from *not* enabling AI

Inteligencia artificial y FDA

Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices

 Email  Print

Software as a Medical Device (SaMD)

[Your Clinical Decision Support Software: Is It a Medical Device?](#)

October 19, 2023 update: 171 Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices were added to the list below. Of those newly added to the list, 155 are devices with final decision dates between August 1, 2022, and July 30, 2023, and 16 are devices from prior periods identified through a refinement of methods used to generate this list.

Content current as of:
12/06/2023

Regulated Product(s)
Medical Devices

2016

Año en el que se aprobó el primer dispositivo

692

Número de *medical devices* de AI aprobados

0

Dispositivos aprobados de IA generativa a octubre 2023

Benjamins, S., Dhunoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ digital medicine, 3(1), 118.

Implementación clínica de las soluciones basadas en aprendizaje automático

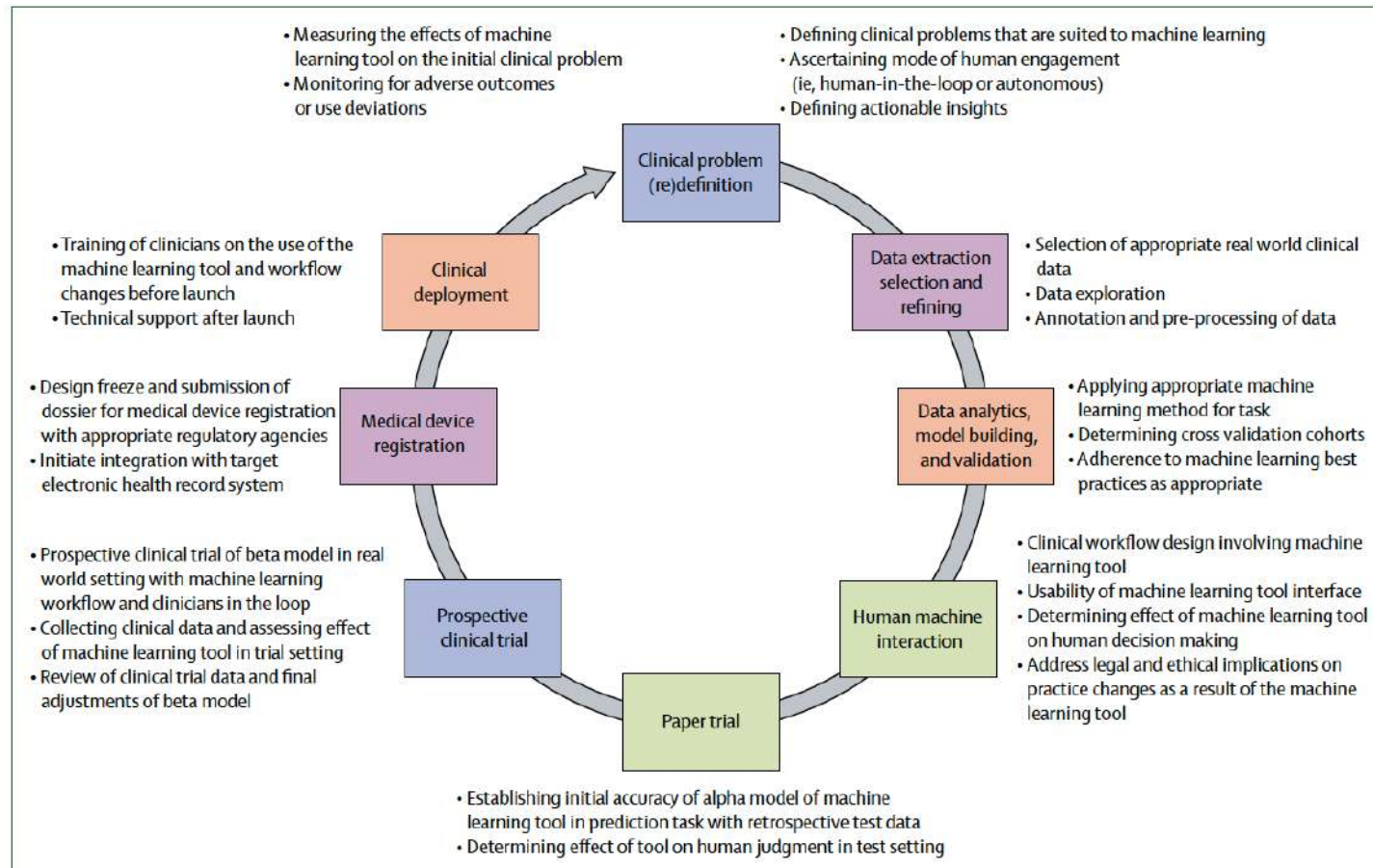


Figure 1: Training, clinical trial evaluation, and clinical implementation of machine learning algorithms for health-care applications

Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262–e273.

Problema ¿Puede ser solucionado con AI?

1. Necesitamos **datos**
2. Necesitamos **permisos** para acceder a los datos
3. Debemos elaborar un **protocolo** y mandarlo al CEIm
4. **Pre-procesar** los datos
5. **Entrenar** modelos
6. **Sesgos, explicabilidad, estabilidad, validación**
7. **Ensayo clínico**
8. **Registrar** el dispositivo
9. Fase de **despliegue**





Introducción a la Inteligencia Artificial

Inteligencia artificial y aprendizaje automático

*Our **intelligence** is what makes us humans, and AI is an extension of that quality – Yann LeCun*

*Artificial **Intelligence** spans a broad set of tools for building machines that exhibit **intelligent** behavior – Andrew NG*

Todo el ML es IA. No toda la IA es ML

Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed – Arthur Samuel

Machine learning
=
aprendizaje
automático

A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E – Tom Mitchell

La inteligencia artificial conoce a la estadística tradicional

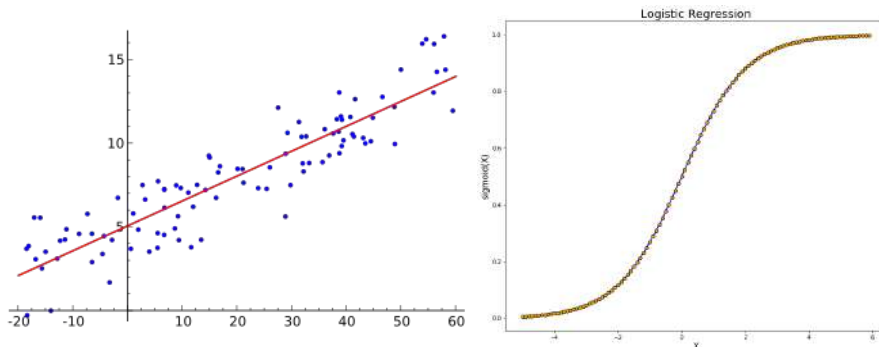
When should linear regression be called "machine learning"?

Asked 6 years, 7 months ago Modified 1 year, 9 months ago Viewed 64k times

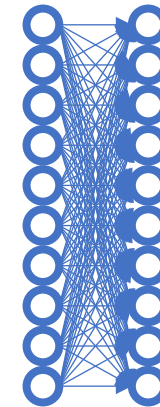
The Two Cultures: statistics vs. machine learning?

Asked 13 years, 3 months ago Modified 2 years, 9 months ago Viewed 175k times

Linear model



Neural network model



*The linear model has been a mainstay of statistics for the past 30 years and remains one of our most important tools – **The Elements of Statistical Learning***

Inteligencia artificial EU AI ACT



Sistema de inteligencia artificial (sistema de IA): el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran debajo y que puede, para un conjunto determinado de objetivos definidos por seres humanos, **generar información de salida como contenidos, predicciones, recomendaciones o decisiones** que influyan en los entornos con los que interactúa.

- Estrategias de **aprendizaje automático**, incluidos el aprendizaje **supervisado**, el no **supervisado** y el realizado por **refuerzo**, que emplean una amplia variedad de métodos, entre ellos el aprendizaje profundo.
- Estrategias basadas en la lógica y el conocimiento, especialmente la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, los sistemas expertos y de razonamiento (simbólico).
- Estrategias **estadísticas**, estimación **bayesiana**, métodos de búsqueda y optimización.



Ratificación de un acuerdo provisional



EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final
2021/0106(COD)

Proposal for a

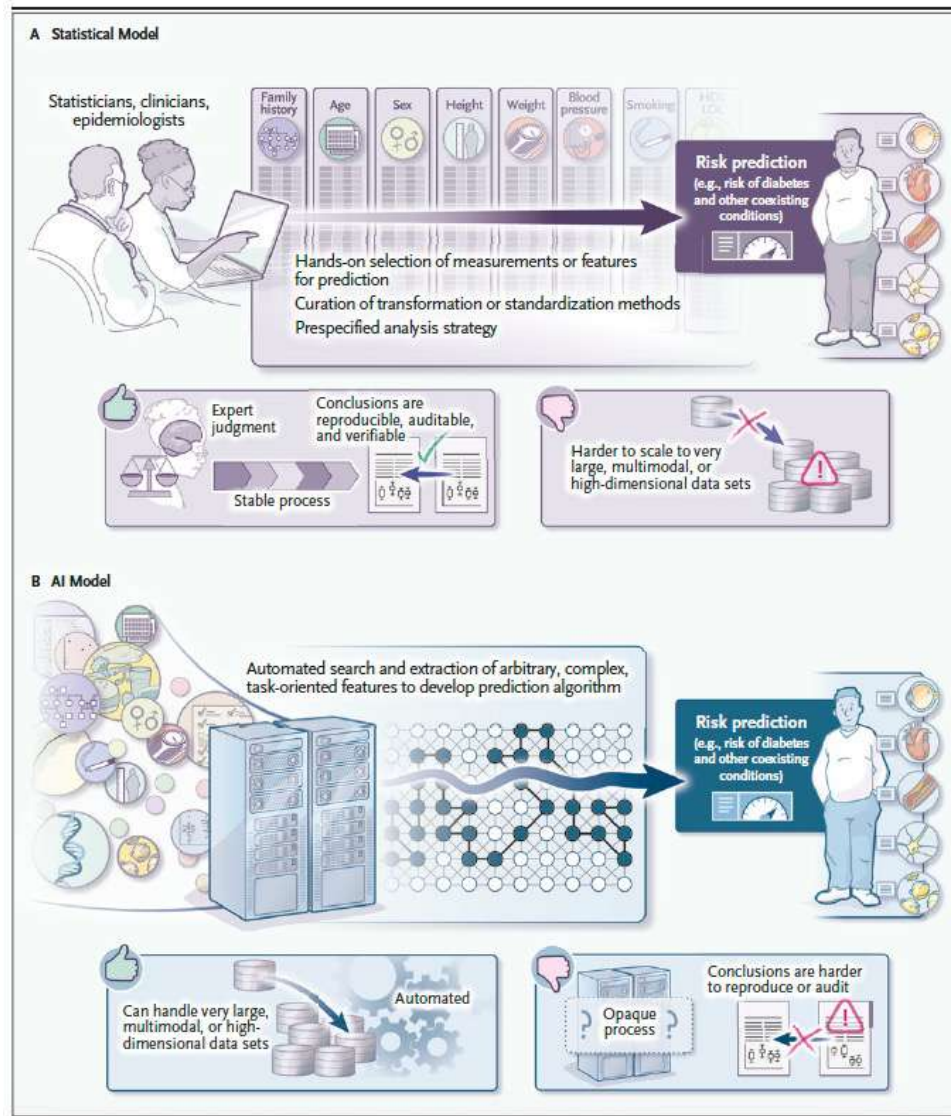
REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

10/11 abril

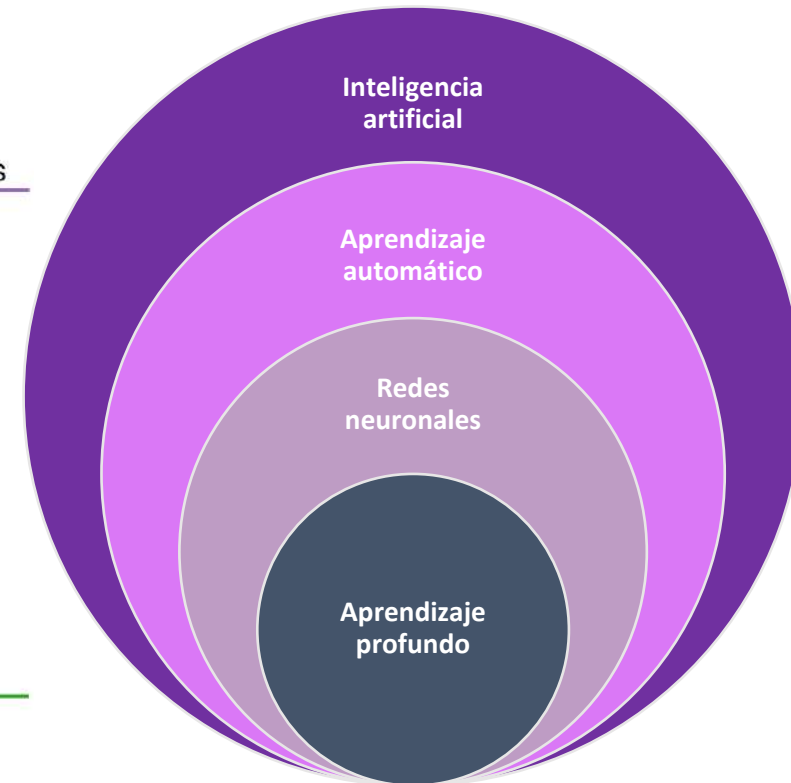
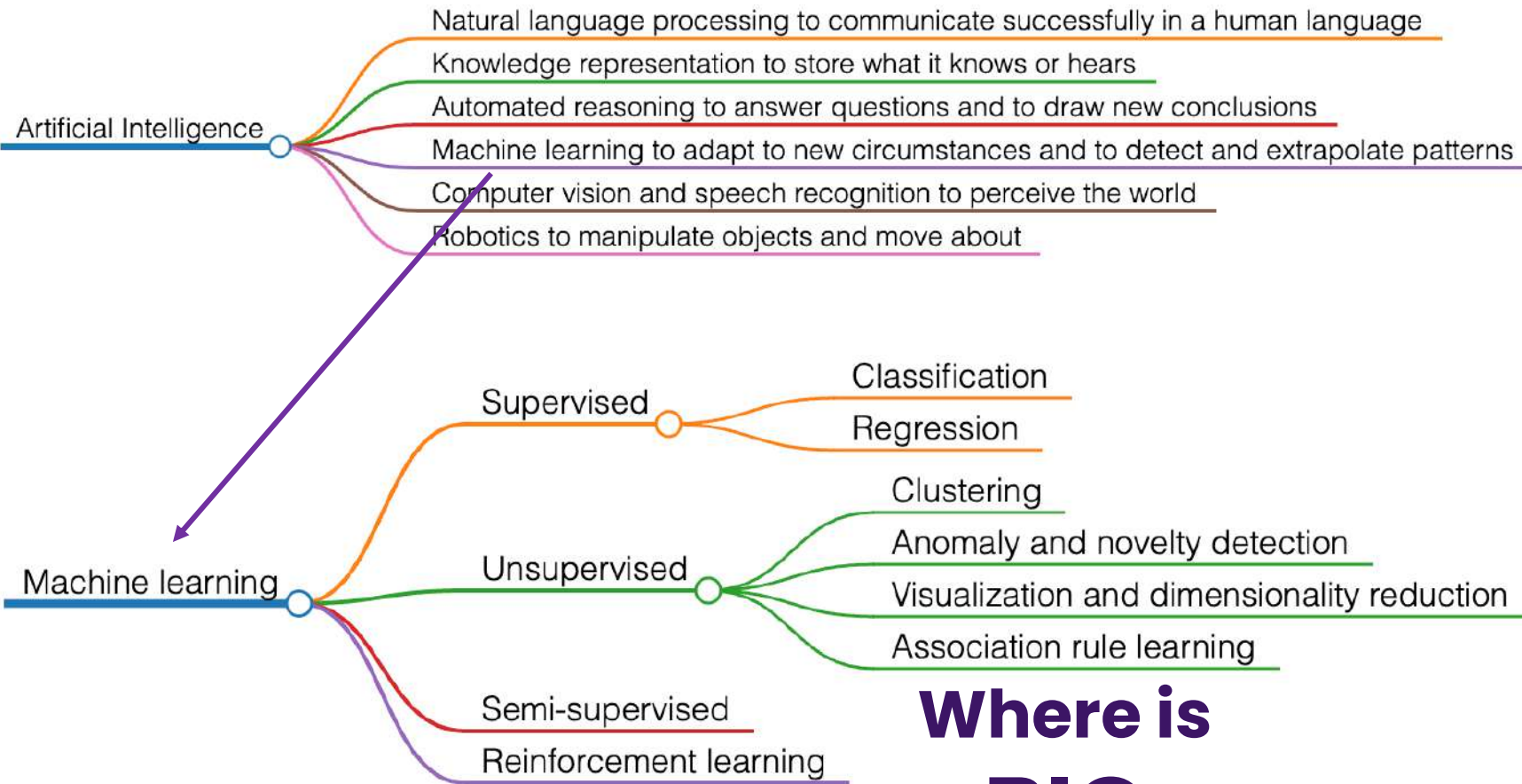
La inteligencia artificial conoce a la estadística tradicional



Estadística tradicional	Inteligencia artificial
Características elegidas manualmente	Manualmente o automáticamente
Conjuntos de datos de no alta-dimensionalidad $n \gg p$	Conjuntos de datos de tamaño variable
Conclusiones reproducibles, auditables y verificables	Conclusiones no tan reproducibles

Hunter, D. J., & Holmes, C. (2023). Where medical statistics meets artificial intelligence. *New England Journal of Medicine*, 389(13), 1211-1219.

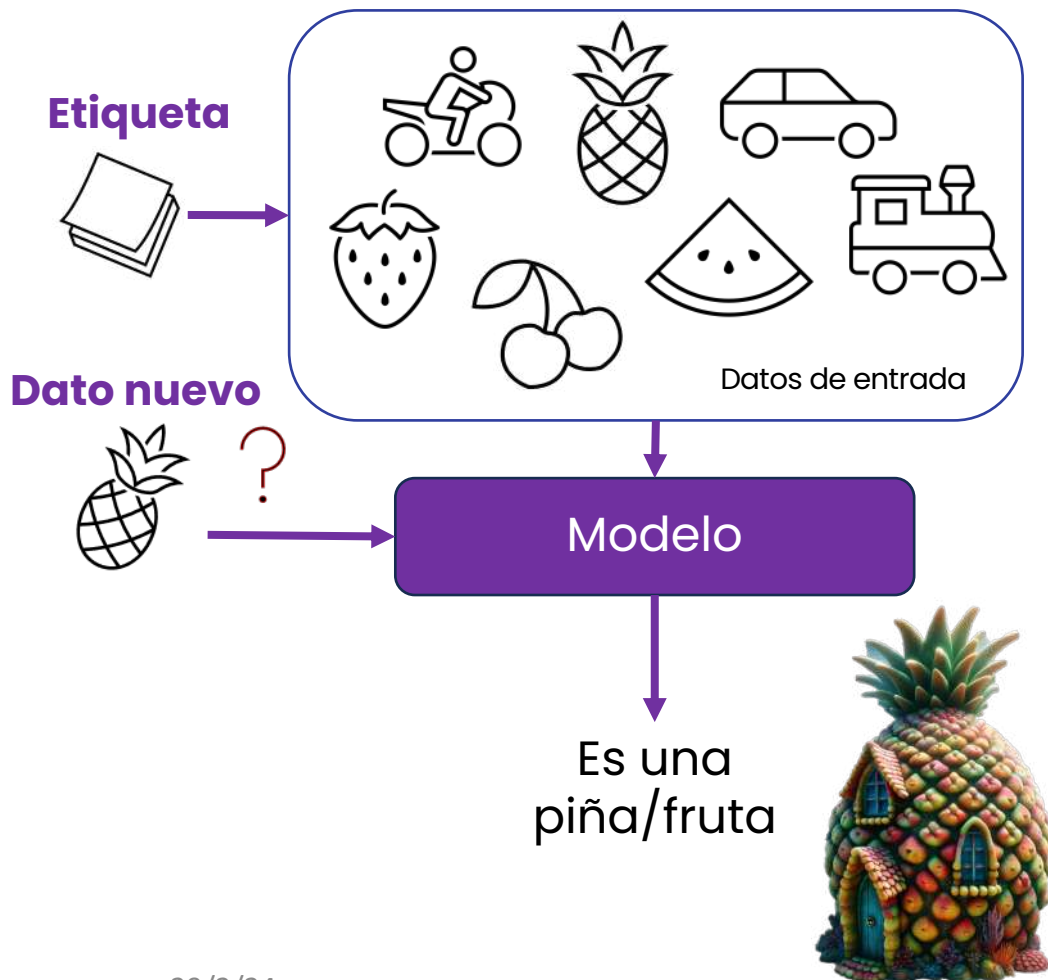
Inteligencia artificial y aprendizaje automático



**Where is
BIG
data?**

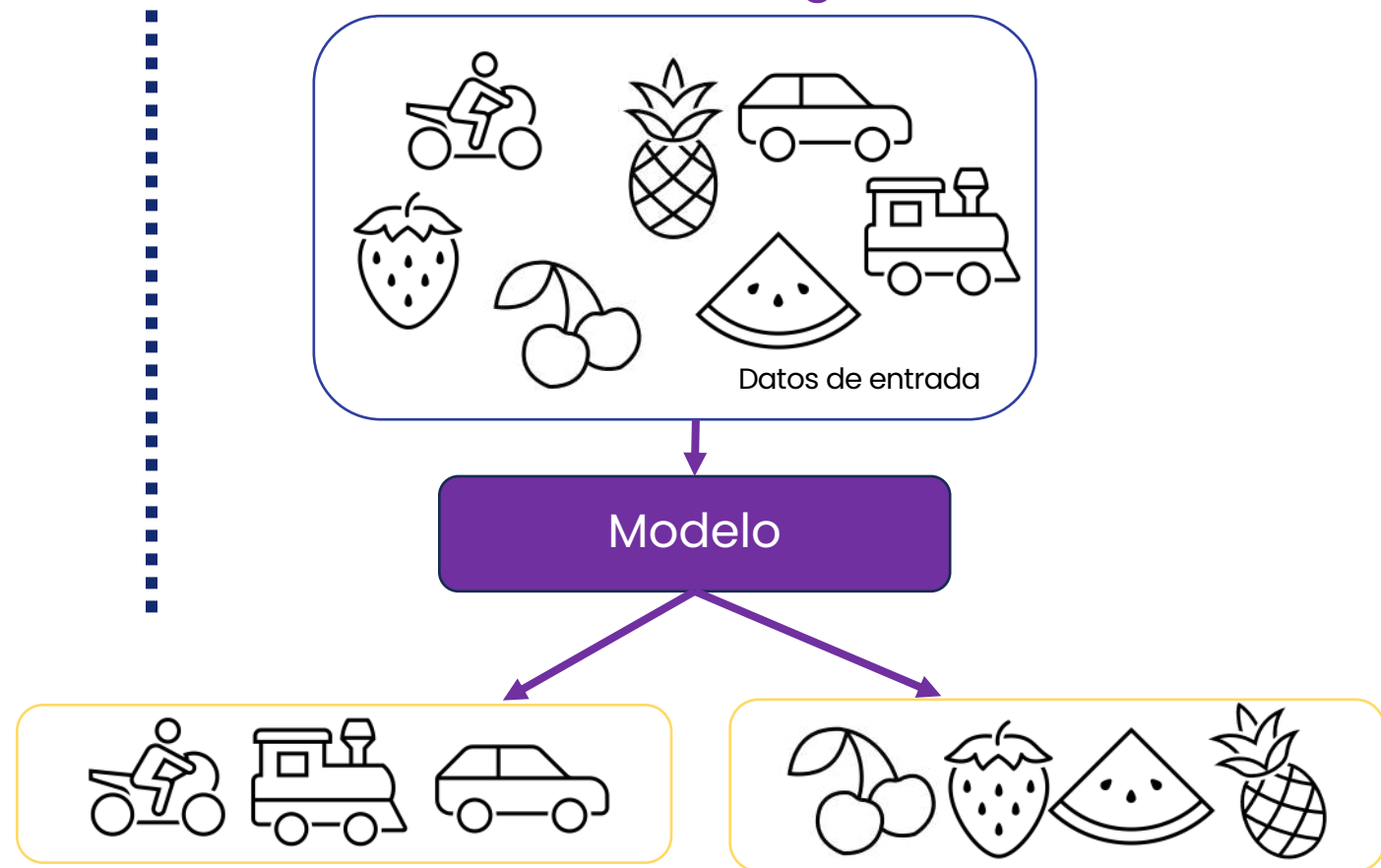
Aprendizaje supervisado

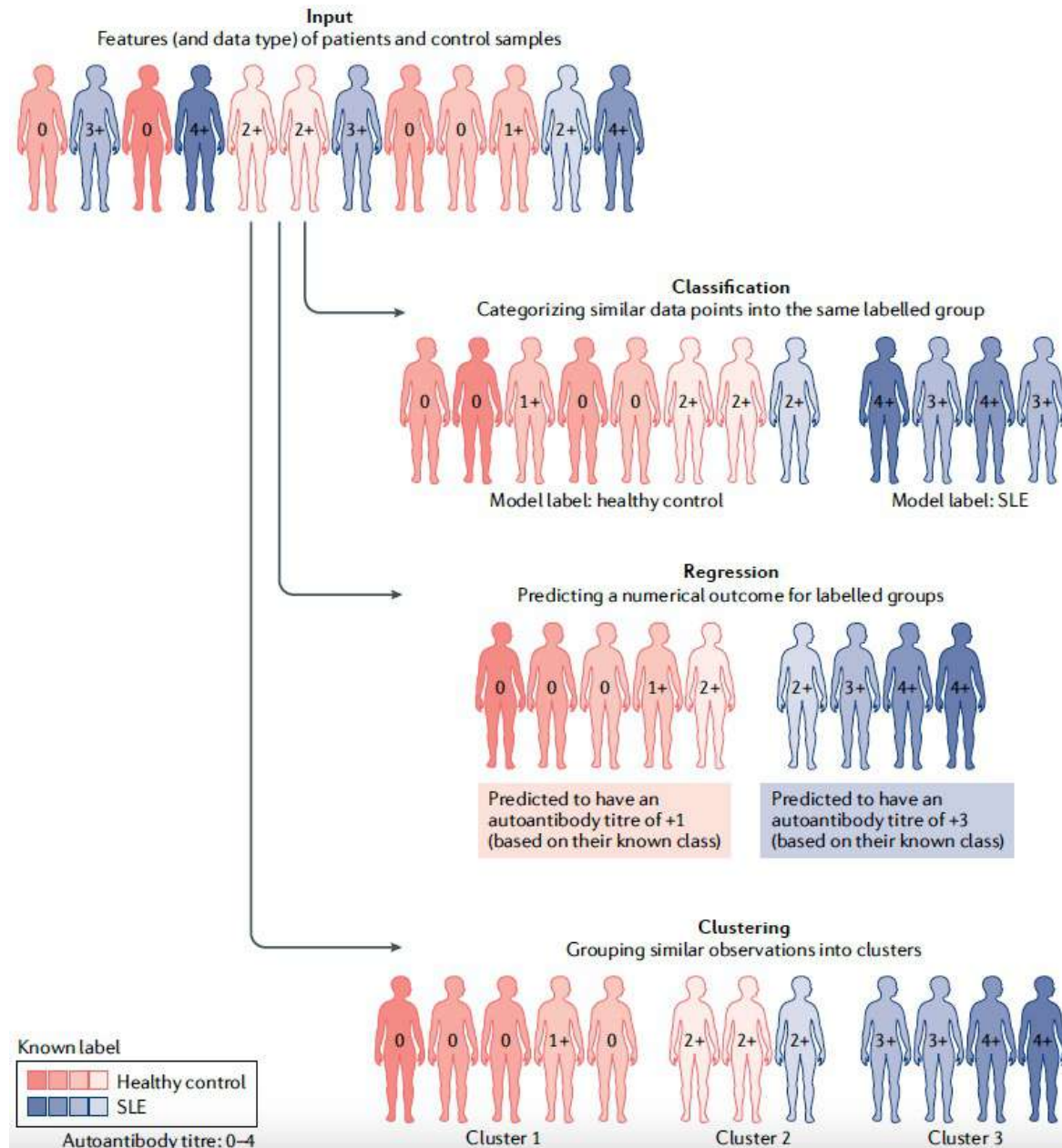
Clasificación



Aprendizaje no supervisado

Clustering





Kingsmore, K. M., Puglisi, C. E., Grammer, A. C., & Lipsky, P. E. (2021). An introduction to machine learning and analysis of its use in rheumatic diseases. *Nature Reviews Rheumatology*, 17(12), 710-730.

Inteligencia artificial y aprendizaje automático



*“If intelligence was a cake, **unsupervised learning** would be the **cake**, **supervised learning** would be the **icing** on the cake, and **reinforcement learning** would be the **cherry** on the cake.” – Yann LeCun*

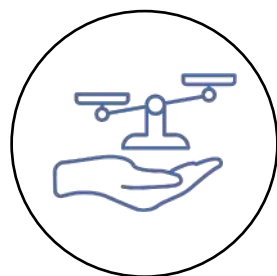
Sobre la importancia de ser capaces de **entrenar modelos** sin necesidad de un conjunto de **datos etiquetado**.

Avances técnicos

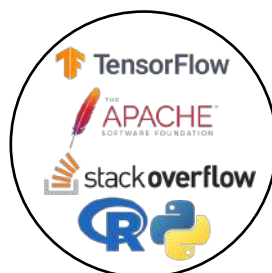
La IA y las técnicas de extracción de datos pueden ofrecer nuevas y prometedoras oportunidades para modelizar y captar interacciones estadísticas complejas

Retos y limitaciones:

hardware, éticas y legales



Herramientas potentes y de **código abierto**, lenguajes **específicos** de ciencia de datos



Plataformas de intercambio de datos, modelos de datos comunes



Infraestructura, marcos, computación distribuida, federada, paralela, nube



0101010
1010101
0101001

Cantidad de **datos** y disponibilidad

Financiación



Aplicación

01

Clasificar enfermedades

02

Predecir enfermedad

03

Identificar predictores

04

Estratificación de pacientes y subgrupos de enfermedad

05

Progresión de la enfermedad y actividad

06

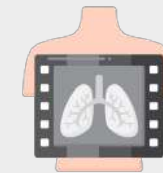
Respuesta al tratamiento



Clinical and demographic



Molecular Biomarkers / genetic



Images

Datos

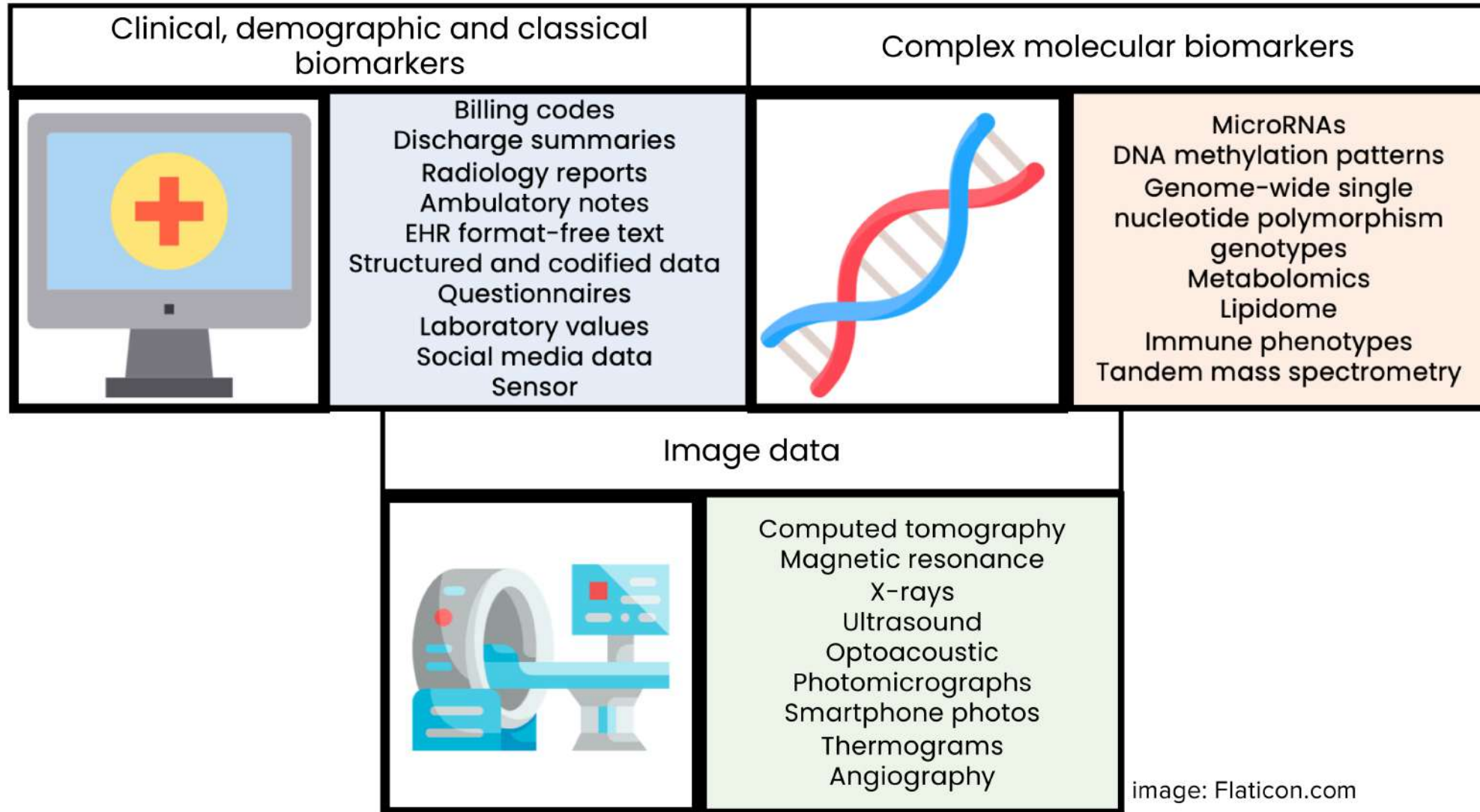




image: Flaticon.com

The background features a complex, abstract pattern of glowing blue and purple lines and dots, resembling a digital or data visualization. The lines are composed of many small, bright points, creating a sense of depth and movement. The overall color palette is dominated by deep blues and purples, with some lighter, shimmering highlights.

**Ejemplo de aplicación de la IA
en enfermedades
musculoesqueléticas**



**RECONOCIMIENTO
DE OCUPACIONES
EN LA EHR DE
REUMATOLOGÍA**

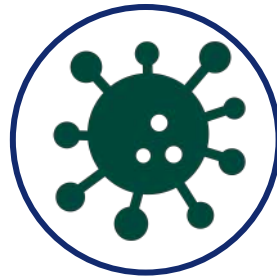
Introducción clínica

La incorporación de información ocupacional en las HCE puede conducir a **diagnósticos clínicos más informados, mejores planes de tratamiento**, políticas más eficaces, intervenciones y estrategias de prevención para mejorar la salud general de la población trabajadora.

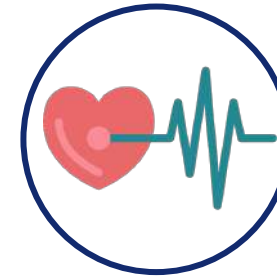
Prevalence: 2 million deaths from work related diseases / year



Risk factors: Asthmagens, carcinogens, ergonomic, risk factors, and noise



Prevention: EU-OSHA, NIOSH



Effect: Mental, physical, inequality, QoL, fertility



Social-Determinants of Health



Not routinely collected in EHR

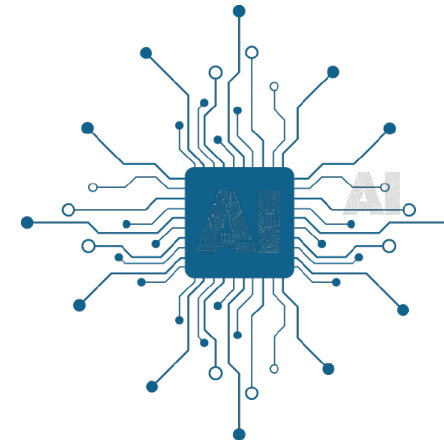
Introducción clínica

1. El estudio de la ocupación como SDoH queda **oculto** por otros determinantes como la raza o la etnia
2. Suele recogerse como **dato no estructurado en texto libre** por lo que es difícil incorporarla a los estudios de investigación
3. En el campo de las enfermedades reumáticas, el capturar información sobre la ocupación es crucial para promover **estrategias de prevención e intervención** para reducir la discapacidad laboral y las bajas por enfermedad

Objetivos

1. Construir un modelo de **aprendizaje profundo** que permita **identificar profesiones** en el texto libre de las historias clínicas electrónicas
2. Evaluar el rendimiento de dichos modelos al extraer menciones de ocupación de pacientes en las **notas clínicas de reumatología**
3. **Describir las características** clínico-demográficas que influyen en la recogida de la ocupación

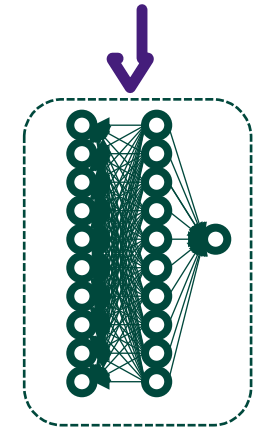
Nota clínica de entrada



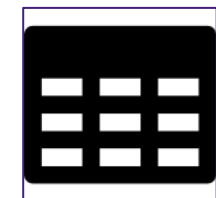
Salida

Objetivo final:
Estructurar

Trabaja como investigador



Trabaja como investigador



Objetivos

- 1 One year after the diagnosis by the physician, the patient experienced significant improvement in his symptoms.
- 2 He reported decreased morning stiffness and a significant reduction in joint pain.
- 3 John had successfully transitioned into a new career as a diving instructor, working as a scuba diving guide.
- 4 Prior to his diagnosis, he stayed in Africa with his brother in law, who worked as a safari guide

**Occupation
Activity
Working status**

**Health professional
Family member
Patient
Other**

**No
occupation**

1	One year after the diagnosis by the physician, the patient experienced significant improvement in his symptoms.	SANITARIO PROFESION	
2	He reported decreased morning stiffness and a significant reduction in joint pain.		
3	John had successfully transitioned into a new career as a diving instructor, working as a scuba diving guide.	PROFESION PACIENTE	PROFESION PACIENTE
4	Prior to his diagnosis, he stayed in Africa with his brother in law, who worked as a safari guide		PROFESION FAMILIAR

Reconocimiento de profesiones en la HCE

Datos de entrenamiento

Aprendizaje Supervisado

Problema de multi-clasificación

Reconocimiento de entidades

← → /TFManotacion/Corpus/S0034-98872012000500010-1

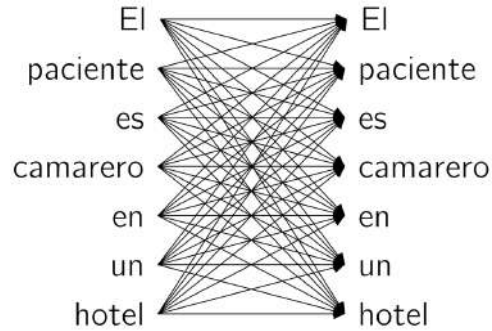
		PROFESION PACIENTE
14	W tenía 35 años, era supervisor de ventas, soltero; sin antecedentes de importancia salvo ser promiscuo; fumaba y bebía poco, no consumía drogas.	
15	En el 2° semestre de 2008 aparecieron tendencia a aislarse, olvidos, insomnio, pérdida de libido, disartria, cefaleas.	
16	Existía una mezcla de labilidad emocional y apatía.	
		SITUACION LABORAL PACIENTE
17	Perdió su trabajo en abril de 2009.	
		SANITARIO PROFESION
		SANITARIO PROFESION
18	En mayo un psiquiatra lo derivó a neurólogo, quien pidió una batería de exámenes, entre ellos VDRL en sangre que fue reactivo.	

Table 2.1: Name entity recognition in MEDDOPROF tasks, represented with BIO schema

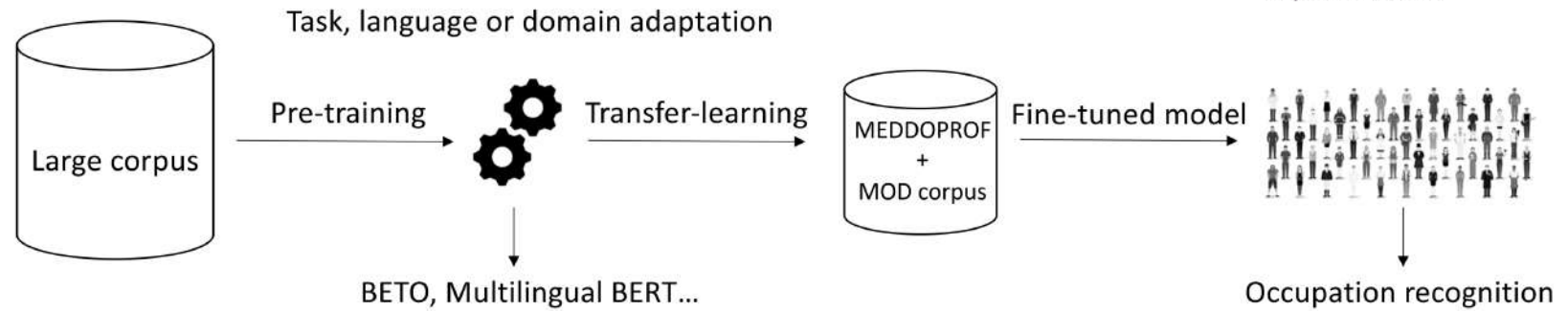
Sentence	El	paciente	es	deportista	profesional	en	activo
Task 1	O	O	O	B-PROFESION	I-PROFESION	O	O
Task 2	O	O	O	B-PACIENTE	I-PACIENTE	O	O

Métodos

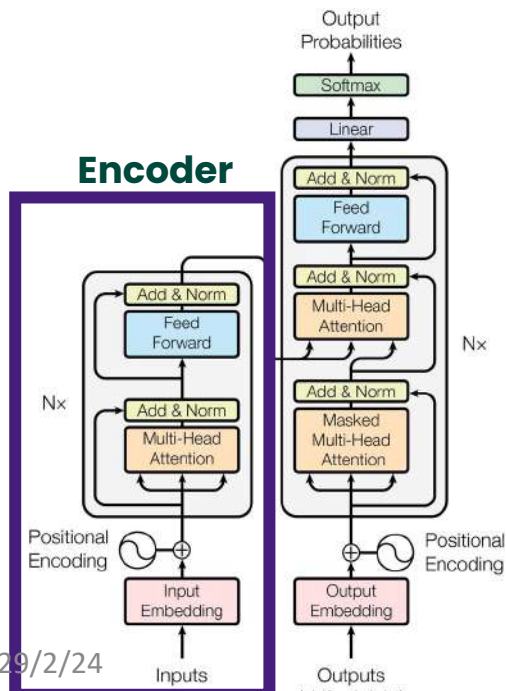
1. Self-attention



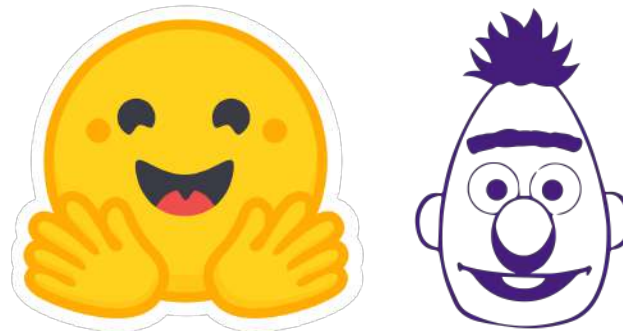
2. Transfer learning



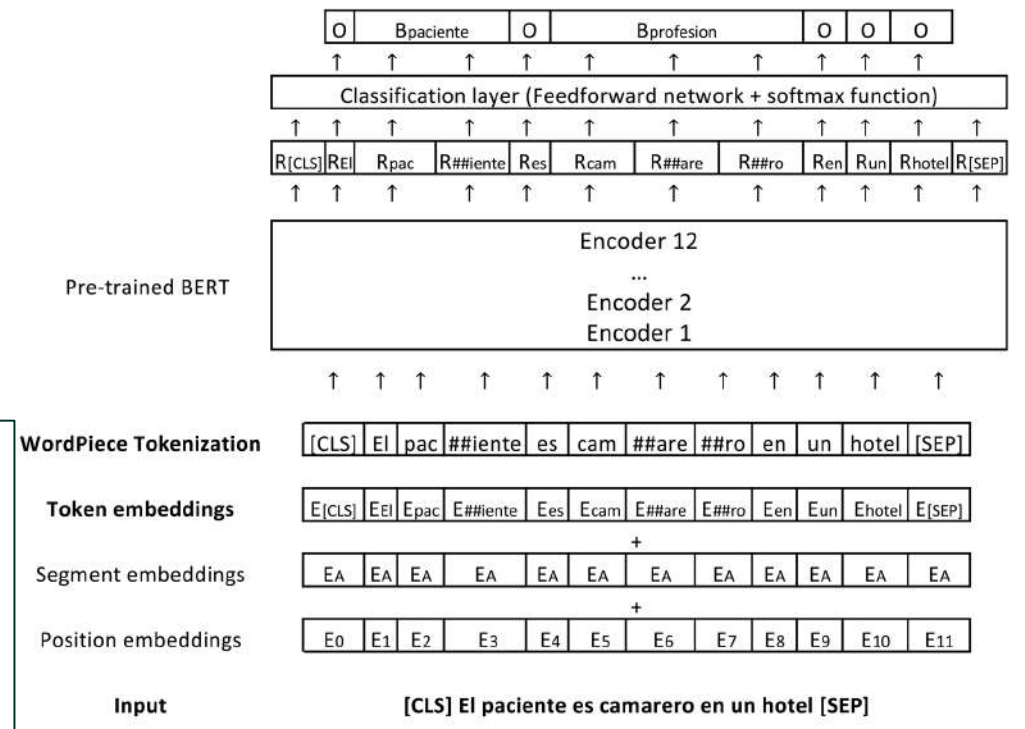
3. Encoder-decoder architecture



4. BERT



Context
Pre-training (MLM, NSP)
WordPiece Tokenizer
Only encoder
Multi-head attention Multiple variants

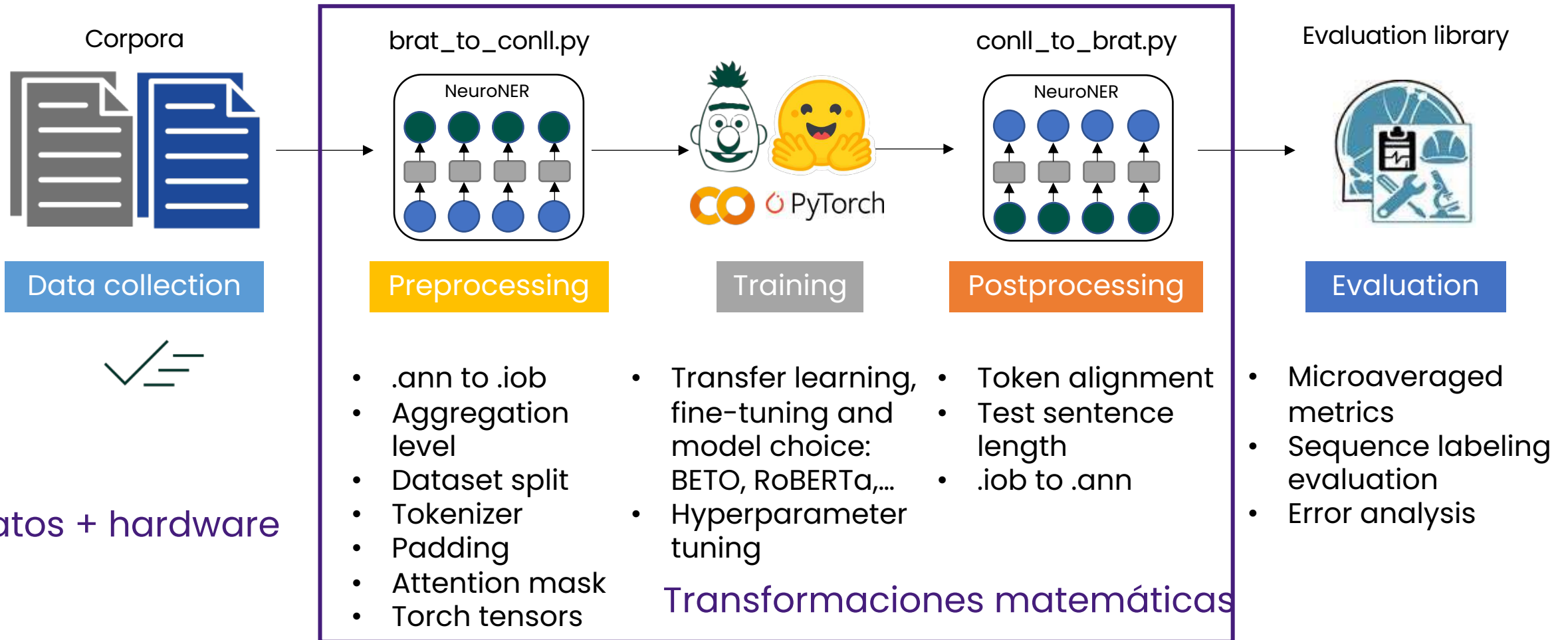


Primer objetivo

1. Construir un modelo de **aprendizaje profundo** que permita **identificar profesiones** en el texto libre de las historias clínicas electrónica

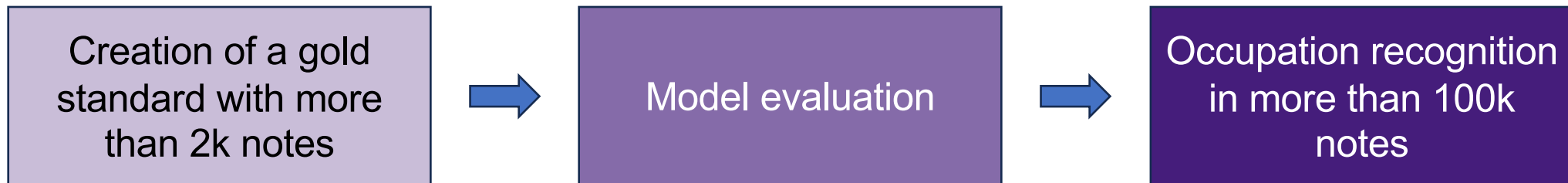
Primer objetivo

RoBERTa: pre-trained with spanish clinical data



Segundo objetivo

2. Evaluar el rendimiento de dichos modelos al extraer menciones de ocupación de pacientes en las **notas clínicas de reumatología**



Reconocimiento de profesiones en la HCE

Bibliotecaria

Trabaja en una cocina

trabaja en colegio con pacientes minusvalidos

Trabaja en un Juzgado de Administrativa

Trabaja en una consulta de Audiología

Tiene una tienda de alimentacion

Cuidadora de un anciano

dependienta de una tienda de alimentacion

interventora del ayuntamiento

Trabaja en una agencia de viajes

Secretaria de Direccion de un banco

Trabaja en un polideportivo

ayudante de realizacion en TV

pintor de brocha gorda

controladora del SER

repartidor de electrodomesticos

Ha trabajado durante anos en quirofano

violinista

profesora de lengua de **senas** en la Universidad

sacerdote

inpectora de sanidad

**Trabaja en la recogida de la fresa
topografo**

preparador fisico de atletismo

trabaja en una Clinica de Radiodiagnostico

empleado en una panificadora

soldador de tela asfaltica

DJ

empleado en el Zoo

monitora de aerobic

Trabajador de mantenimiento en empresa aerea

reponedora en un hipermercado

planchadora hace 26 anos en tintoreria

Trabaja mucho ocn las manos de pinche de cocina

trabaja desmontando pescando y productos de
ultracongelados

**Trabaja en montaje de electricas en torres
de alta tension**

psicologa en un centrodde ninos discapacitados

conductor de autobuses de la EMT

Vigilante de seguridad

empleada en cruz roja

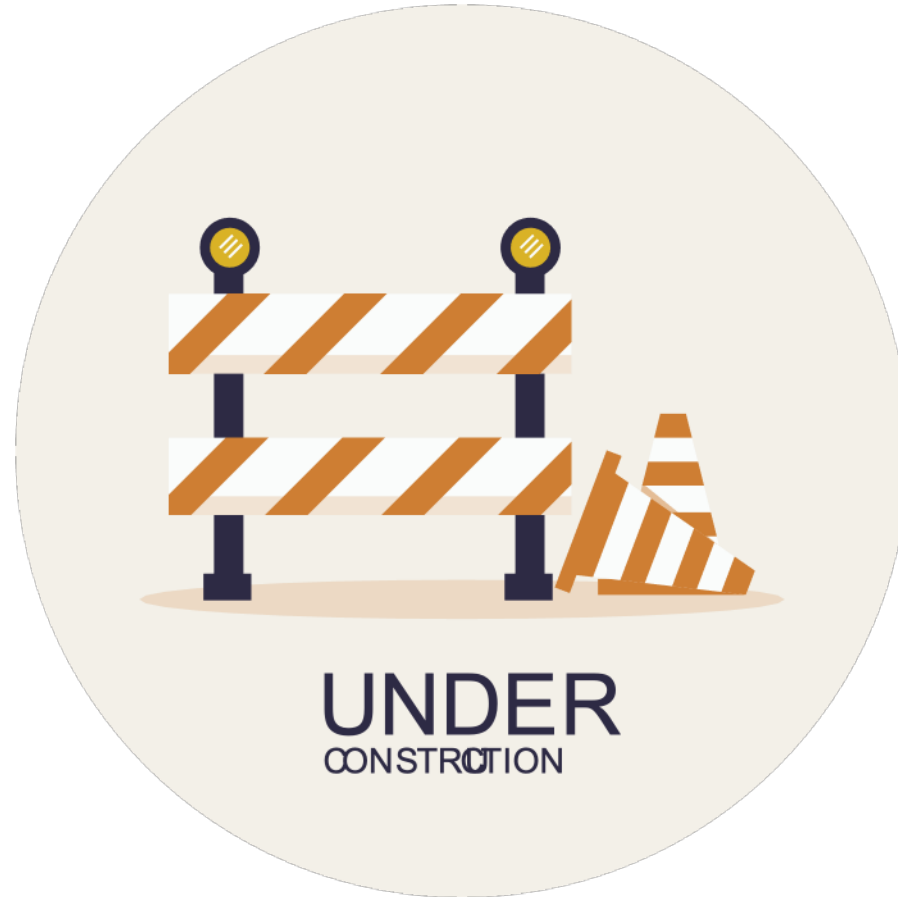
**trabaja en el aeropuerto de Barajas
maneja martillos neumaticos**

Tercer objetivo

3. Describir las características clínico-demográficas que influyen en la recogida de la ocupación



Después de realizar los análisis bivariados y multivariados vimos que los pacientes con diagnósticos relacionados con **patología mecánica altamente incapacitante** (es decir, dolor de espalda, trastornos musculares) se asociaron con una mayor probabilidad de recopilación de ocupación.





OTRAS CUESTIONES

Retos principales

Algoritmos

1. Cantidad insuficiente de datos de entrenamiento

2. Datos de entrenamiento no representativos

3. Pobre calidad de los datos

4. Características poco relevantes

5. Sobreentrenamiento

6. Subajuste

Éticas

1. Selección del problema

2. Recogida de datos

Otras

1. Regulación

2. Falta de estandarización

3. Validación externa

4. Explicabilidad

Tenemos muchos datos, datos interesantes, datos sensibles, datos de distintos tipos, pero si no son de calidad ...

...el producto final no va a ser útil

Calidad de los datos

Tabla 1 - Características de calidad de datos definidas en la Norma UNE-ISO/IEC 25012 [6]

Característica UNE-ISO/IEC 25012	Inherente	Dependiente del Sistema
Exactitud	X	
Compleitud	X	
Consistencia	X	
Credibilidad	X	
Actualidad	X	
Accesibilidad	X	X
Conformidad	X	X
Confidencialidad	X	X
Eficiencia	X	X
Precisión	X	X
Trazabilidad	X	X
Comprensibilidad	X	X
Disponibilidad		X
Portabilidad		X
Recuperabilidad		X



most of today's knowledge on disease mechanisms **will be forgotten** and we will be ruled by systems that only focus on

int
ou
Th
co
AI

Viewpoint | AI in Medicine

FREE

January 19, 2024

AI's Threat to the Medical Profession

Agnes B. Fogo, MD¹; Andreas Kronbichler, MD, PhD²; Ingeborg M. Bajema, MD, PhD³

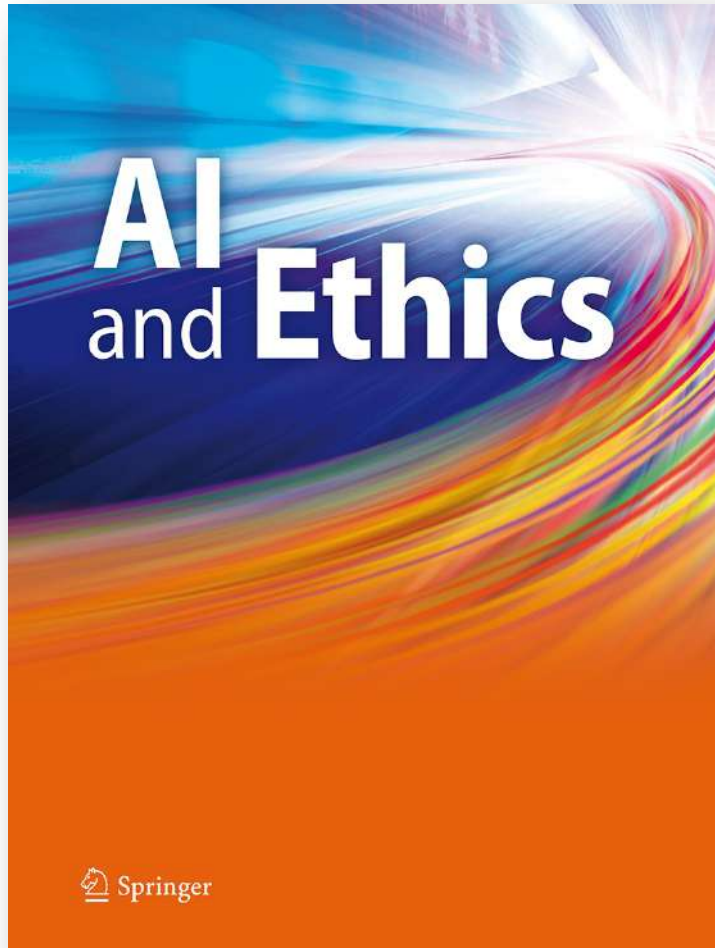
» [Author Affiliations](#) | [Article Information](#)

JAMA. 2024;331(6):471-472. doi:10.1001/jama.2024.0018

because there is **no alternative expertise** to form the basis for any debate.

Fogo AB, Kronbichler A, Bajema IM. AI's Threat to the Medical Profession. *JAMA*. 2024;331(6):471-472. doi:10.1001/jama.2024.0018

Aprendizaje automático y ética en salud



AI and Ethics seeks to promote informed debate and discussion of the ethical, regulatory, and policy implications that arise from the development of AI

Iniesta, R. (2023). The human role to guarantee an ethical AI in healthcare: a five-facts approach. *AI and Ethics*, 1-13.

Inteligencia artificial y ética en salud



“We describe ongoing efforts and outline challenges in a proposed pipeline of ethical ML in health, ranging from problem selection to postdeployment considerations. We close by summarizing recommendations to address these challenges.”

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical machine learning in healthcare. *Annual review of biomedical data science*, 4, 123-144.

the **greatest danger would be AI that's implemented by players that don't necessarily have the patient's best interests at heart** Isaac Kohane

Will Patients Have To **Pay** For Using AI In Their Healthcare?

Modelos justos, fiables y explicables

arXiv:1802.04422v1 [stat.ML] 13 Feb 2018

A comparative study of fairness-enhancing interventions in machine learning *

Sorelle A. Friedler, Haverford College[†]
Carlos Scheidegger, University of Arizona[‡]
Suresh Venkatasubramanian, University of Utah[§]
Sonam Choudhary, University of Utah[¶]
Evan P. Hamilton, Haverford College^{||}
Derek Roth, Haverford College^{**}

Abstract

Computers are increasingly used to make decisions that have significant impact in people's lives. Often, these predictions can affect different population subgroups disproportionately. As a result, the issue of fairness has received much recent interest, and a number of fairness-enhanced classifiers and predictors have appeared in the literature. This paper seeks to study the following questions: how do these different techniques fundamentally compare to one another, and what accounts for the differences? Specifically, we seek to bring attention to many under-appreciated aspects of such fairness-enhancing interventions. Concretely, we present the results of an open benchmark we have developed that lets us compare a number of different algorithms under a variety of fairness measures, and a large number of existing datasets. We find that although different algorithms tend to prefer specific formulations of fairness preservations, many of these measures strongly correlate with one another. In addition, we find that fairness-preserving algorithms tend to be sensitive to fluctuations in dataset composition (simulated in our benchmark by varying training-test splits), indicating that fairness interventions might be more brittle than previously thought.

*This work was partially supported by National Science Foundation under grants IIS-1633387, IIS-1513651, and IIS-1633724, as well as by a grant from the Ethics and Governance of AI Initiative. Source code and instructions for adding your own algorithm or dataset, can be found at: <https://github.com/aigofairness/fairness-comparison>
[†]sorelle@cs.haverford.edu
[‡]carlos@cs.arizona.edu
[§]suresh@cs.utah.edu
[¶]sonam@cs.utah.edu
^{||}evanphamilton@gmail.com
^{**}derek.roth1@gmail.com

CONTRIBUTED RESEARCH ARTICLE

fairmodels: a Flexible Tool for Bias Detection, Visualization, and Mitigation in Binary Classification Models

Jakub Winiński, *Preprint* <https://arxiv.org/abs/2005.07072>

Abstract Machine learning decision systems are becoming omnipresent in our lives. From dating apps to voting, loan seeking, algorithm affect both our well-being and future. Typically, however, these systems are not auditable. Moreover, complex predictive models are eager to learn social biases present in historical data that can lead to increasing discrimination. It is not so easy to make especially these models and tools to do deep validation of models also from the perspective of potential discrimination. This work introduces an R package *fairmodels* that helps to visualize fairness and eliminate bias in binary classification models easily and flexibly. The *fairmodels* package offers a model-agnostic approach to bias detection, visualization, and mitigation. The implemented set of functions and fairness metrics enable model fairness validation from different perspectives. The package includes a suite of methods for bias mitigation that aim to eliminate the discrimination in the model. The package is designed not only to assess a single model but also to facilitate comparisons between multiple models.

Introduction

Responsible machine learning and in particular fairness are gaining attention within a machine learning community. The reason for this is that predictive algorithms are becoming more and more decision and influential in our lives. This impact could be low or non-significant and in some ranging from user's feeds on social platforms, displayed ads, and recommendations on an online store to loan decisions, social scoring, and credit recognition systems used by banks and authorities. Because it leads to automated systems that learn some undesired bias preserved in data for socio-historical reasons. What is making a job (Loh et al., 2019) or hiring one's data processed by credit systems (Lopez et al., 2016), sensitive attributes such as race, religion, ethnicity, etc. might play a major role in the decision. Even if such variables are not directly included in the model, they are often captured by proxy variables such as zip code to proxy for race and wealth, purchase products to proxy for gender and age, eye color to proxy for ethnicity. As one would expect they can give an unfair advantage to a particular group. Discrimination takes the form of more favorable predictions or higher accuracy for a particular group. For example, some popular commercial gender classifiers were found to perform the worst on darker-skinned (Bastien and Larue, 2017). From now on, such social and historical attributes towards people with specific sensitive attributes will be called bias.

The list of personal attributes may depend on the domain for which the model is built as well as on the country. For example, the European Union law is contained in the Handbook on European non-discrimination law (European Union Agency for Fundamental Rights and Council of Europe, 2015), which lists the following protected attributes that cannot be the basis for making treatment: sex, gender identity, sexual orientation, disability, age, race, ethnicity, nationality, racial or ethnic origin, religion or belief, social origin, birth, and property, language, political or other opinions. This list, though long, does not include all potentially relevant items, e.g. in the USA, a protected attribute is the pregnancy, the status of a nurse, or genetic information.

While there are historical and economical reasons for this to happen, such decisions are unacceptable for ethical reasons and sometimes are prohibited by local laws. The problem is not simple, especially when the only criterion set for the system is performance. In some problems, we observe a trade-off between accuracy and fairness to less bias discrimination, leads to lower performance (Lopez and Larue, 2017). Sometimes, bias, which are considered ground truth might also be biased (Fink et al., 2017) and when controlling for that bias the performance and fairness might improve at the same time. Most of the time however what we want to improve, fairness from one perspective or becomes some or another (Barnes et al., 2019).

The bias in machine learning systems has potentially many different sources. In *Statistics of (D)I* authors categorized their sites on types like: historical bias, where information is already collected into the data reflecting the world, observer bias, sampling bias, ranking and social biases, and many more. That shows how many designs are potentially hidden in the data itself. Whether one would like to act on it or not, it is essential to detect bias and make well-informed decisions whose consequences could potentially harm every group of people. Representations of such systems can be un-predictable. As argued by (Barnes et al., 2017) machine learning systems are not able to aggregate the dispersion.

The R Journal Vol. XX/YY, AAAA 2022

ISSN 2073-4859

Statistics and Computing (2022) 32:7
<https://doi.org/10.1007/s11222-022-10143-w>

Achieving fairness with a simple ridge penalty

Marco Scutari¹, Francesca Panero², Manuel Prüss³

Received: 16 August 2021 / Accepted: 31 August 2022 / Published online: 18 September 2022
© The Author(s) 2022

Abstract

In this paper, we present a general framework for estimating regression models subject to a user-defined level of fairness. We enforce fairness as a model selection step in which we choose the value of a ridge penalty to control the effect of sensitive attributes. We then estimate the parameters of the model conditional on the chosen penalty value. Our proposal is mathematically simple, with a solution that is partly in closed form and produces estimates of the regression coefficients that are linear in the intercept as a function of the level of fairness. Furthermore, it is easily extended to generalized linear models, kernelized regression models and other penalties, and it can accommodate multiple definitions of fairness. We compare our approach with the regression model from Koyama et al. (in: Proceedings of machine learning research, 35th international conference on machine learning (ICML), vol 80, pp 2737–2746, 2018), which implements a proximal optimal linear regression model and with the fair models from Zafar et al. (J Mach Learn Res 20:1–42, 2019). We evaluate these approaches empirically on six different data sets, and we find that our proposal provides better goodness of fit and better predictive accuracy for the same level of fairness. In addition, we highlight a source of bias in the original experimental evaluation in Koyama et al. (in: Proceedings of machine learning research, 35th international conference on machine learning (ICML), vol 80, pp 2737–2746, 2018).

Keywords Linear regression · Logistic regression · Generalised linear models · Fairness · Ridge regression

1 Introduction

Machine learning models are increasingly being used in applications where it is crucial to ensure the accountability and fairness of the decisions made on the basis of their output: some examples are criminal justice (Berk et al., 2021), credit risk modelling (Fuster et al., 2020) and screening job applicants (Haghtan et al., 2020). In such cases, we are required to ensure that we are not discriminating individuals based on sensitive attributes such as gender and race, lead-

ing to disparate treatment of specific groups. At the same time, we would like to achieve the best possible predictive accuracy over other predictors.

The task of defining a non-discriminating treatment, though, does not come without challenges. The concept of fairness itself, in fact, has been characterized in different ways depending on the context. From an ethical and legal perspective, for example, it might depend on the type of distortion we wish to limit, which in turn works with the type of application. Sometimes, we want to limit the adverse bias against a specific group, while in other instances we wish to protect single individuals. Alongside the legal and philosophical research debate, institutional regulations on the use of algorithms in society have been proposed in the last decade. For a comparison among the USA, EU and UK regulations, see Cahn et al. (2018). The European Commission has recently released the first legal framework for the use of artificial intelligence (European Commission 2021), which is now under revision by the member states.

At the same time, there has been a growing interest towards fairness-preserving methods in the machine learning literature. From a statistical perspective, different charac-

[†] Marco Scutari

[‡] Francesca Panero

[§] Manuel Prüss

^{||} manuel.pruss@ethz.ch

[¶] manuel.pruss@ethz.ch

^{||} manuel.pruss@ethz.ch

^{**} manuel.pruss@ethz.ch

^{††} manuel.pruss@ethz.ch

^{‡‡} manuel.pruss@ethz.ch

^{§§} manuel.pruss@ethz.ch

^{¶¶} manuel.pruss@ethz.ch

^{|||} manuel.pruss@ethz.ch

fairness

fairmodels

fairml

Corregir sesgos algorítmicos (como el lenguaje, el género o los políticos) en los procesos de decisión automatizados basados en modelos de aprendizaje automático

Estabilidad de los modelos

Riley et al. *BMC Medicine* (2023) 21:502
https://doi.org/10.1186/s12916-023-03212-y

BMC Medicine

OPINION Open Access

Clinical prediction models and the multiverse of madness

Richard D. Riley^{1,2*}, Alexander Pate³, Paula Dhirman⁴, Lucinda Archer^{1,2}, Glen P. Martin³ and Gary S. Collins⁴

Abstract
Background Each year, thousands of clinical prediction models are developed to make predictions (e.g. estimated risk) to inform individual diagnosis and prognosis in healthcare. However, most are not reliable for use in clinical practice.
Main body We discuss how the creation of a prediction model (e.g. using regression or machine learning methods) is dependent on the sample and size of data used to develop it—were a different sample of the same size used from the same overarching population, the developed model could be very different even when the same model development methods are used. In other words, for each model created, there exists a multiverse of other potential models for that sample size and, crucially, an individual's predicted value (e.g. estimated risk) may vary greatly across this multiverse. The more an individual's prediction varies across the multiverse, the greater the instability. We show how small development datasets lead to more different models in the multiverse, often with vastly unstable individual predictions, and explain how this can be exposed by using bootstrapping and presenting instability plots. We recommend healthcare researchers seek to use large model development datasets to reduce instability concerns. This is especially important to ensure reliability across subgroups and improve model fairness in practice.
Conclusions Instability is concerning as an individual's predicted value is used to guide their counselling, resource prioritisation, and clinical decision making. If different samples lead to different models with very different predictions for the same individual, then this should cast doubt into using a particular model for that individual. Therefore, visualising, quantifying and reporting the instability in individual-level predictions is essential when proposing a new model.
Keywords Clinical prediction model, Instability, Variance, Risk prediction, Bootstrapping, Mean absolute prediction error (MAPE)

Background
The multiverse refers to the potentially infinite number of other universes besides our own, which may or may not be similar. Related concepts are multiple realities, parallel worlds, and alternate universes. Although the multiverse is hypothetical, it gains growing popularity in science-fiction novels and films such as *Spider-Man: Into the Spider-Verse* and *Doctor Strange and the Multiverse of Madness*. The idea of an infinite number of different realities is reflected in the theory of probability and statistics, which acknowledges the variability across random samples of the same size taken from a particular population.

*Correspondence: Richard D. Riley, rd.riley@bham.ac.uk
¹ College of Medical and Dental Sciences, Institute of Applied Health Research, University of Birmingham, Birmingham B15 2TT, UK
² National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, Birmingham, UK
³ Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
⁴ Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7JD, UK

BMC

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

most are not reliable for use in clinical practice

“Whenever researchers develop a clinical prediction model using a particular dataset and model development approach, we recommend they use bootstrapping to investigate the corresponding multiverse of models and instability of individual predictions”

Riley, R.D., Pate, A., Dhirman, P. et al. Clinical prediction models and the multiverse of madness. *BMC Med* 21, 502 (2023). <https://doi.org/10.1186/s12916-023-03212-y>

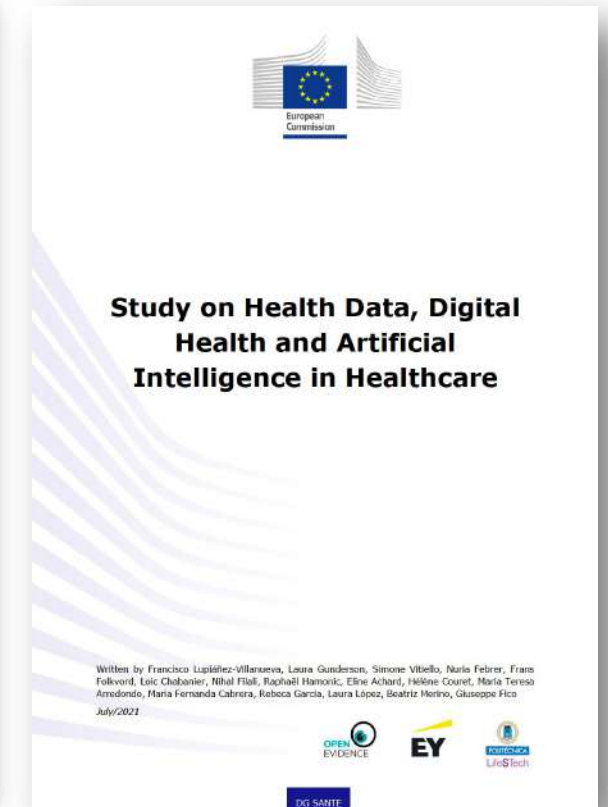
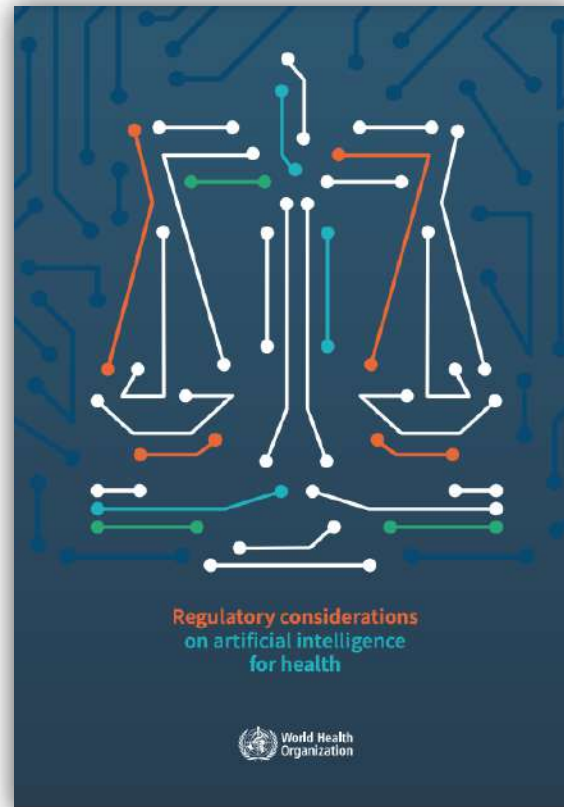
Guías de reporte



Klontzas, M. E., Gatti, A. A., Tejani, A. S., & Kahn Jr, C. E. (2023). AI Reporting Guidelines: How to Select the Best One for Your Research. *Radiology: Artificial Intelligence*, 5(3), e230093.

Acrónimo	Objetivo	Formato
CLAIM	Applications in medical imaging	Checklist
CONSORT-AI	Clinical trials involving interventions	Checklist
SPIRIT-AI	Clinical trials involving interventions	Checklist
FUTURE-AI	Safe, trusted, and ethical solutions for medical imaging	Set of guiding principles, checklist
MI-CLAIM	Minimum information guidelines for applications in medicine	Process map with parts, steps
MINIMAR	Minimum information guidelines for reporting classification and prediction models in health care	List of guiding/essential principles with sub-components
PROBAST-AI*	Asses risks of bias and concerns of applicability prediction model studies	Four steps detailing specific tasks and timeline for completion
RQS	Assess quality of radiomics applications	A set of 16 criteria summarizing key components of radiomics studies
STARD-AI*	Standard for reports of diagnostic accuracy evaluation model studies	Checklist
TRIPOD-AI*	Standard for reports of diagnostic accuracy evaluation model studies	Checklist

Nuevos marcos regulatorios



Shaping Europe's digital future
<https://digital-strategy.ec.europa.eu/en/library>

Publications Office of the European Union:
<https://op.europa.eu/en/home>



¿Cómo
estar al día
con los
últimos
avances en
IA y ML?

¿Cómo estar al día con los últimos avances de IA?



<https://twitter.com/rasbt>
<https://twitter.com/ylecun>
<https://twitter.com/AndrewYNg>
<https://twitter.com/fchollet>
https://twitter.com/goodfellow_ian
<https://twitter.com/akhaliq>
<https://twitter.com/geoffreyhinton>



<https://www.reddit.com/r/MachineLearning>
<https://www.reddit.com/r/LanguageTechnology/>
<https://www.reddit.com/r/learnmachinelearning>
<https://www.reddit.com/r/artificial/>



<https://paperswithcode.com/>
<https://huggingface.co/papers>
<https://ai.meta.com/blog/>
<https://blog.research.google/>
<https://pydata.org/upcoming-events/>
<https://www.linkedin.com/in/omarsar/>

¿Cómo estar al día con los últimos avances de IA?



BIOMED

Sociedad Española de
Inteligencia Artificial en Biomedicina



ASOCIACIÓN
SALUD DIGITAL

La (R)evolución de la IA, en la Salud

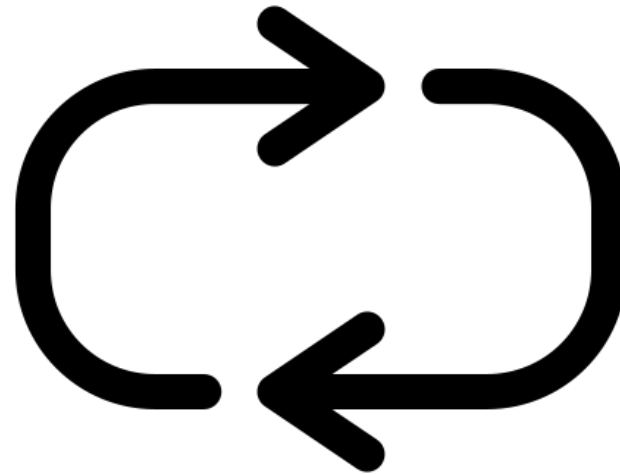


PUNTOS PARA LLEVAR A CASA

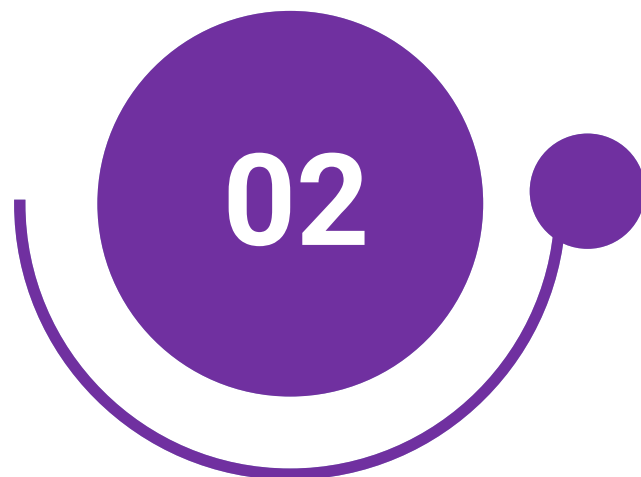
Puntos para llevarse a casa



Muchos avances, pero también mucha incertidumbre



Puntos para llevarse a casa



La adopción es una realidad

AI Adoption in U.S. Health Care Won't Be Easy

by James B. Rebitzer and Robert S. Rebitzer

September 14, 2023

RESEARCH

Why is AI adoption in health care lagging?

Avi Goldfarb and Florenta Teodoridis

March 9, 2022

Preguntas





IdISSC

INSTITUTO de
INVESTIGACIÓN SANITARIA
Hospital Clínico San Carlos

ÓN SANITARIA
an Carlos

Papel y aplicación de las técnicas de inteligencia artificial en medicina

Alfredo Madrid García, PhD

BACK-UP SLIDES

Revistas científicas