

III Jornada de Innovación en Cuidados Continuos

MADRID, 4 DE DICIEMBRE 2025



Integración de la tecnología en la práctica clínica: APPs y dispositivos electrónicos

Juan Cristóbal Sánchez González

Servicio de Oncología Médica

Hospital Universitario Puerta de Hierro Majadahonda

AGENDA

1.- TECNOLOGÍA Y CLÍNICA

2.- EJEMPLOS Y VALIDACIÓN

3.- BIG-DATA, CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

4.- INTEGRAR LA TECNOLOGÍA

5.- CONCLUSIONES

AGENDA

1.- TECNOLOGÍA Y CLÍNICA

2.- EJEMPLOS Y VALIDACIÓN

3.- BIG-DATA, CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

4.- INTEGRAR LA TECNOLOGÍA

5.- CONCLUSIONES

El avance tecnológico en los últimos años supera la capacidad de adaptación del sistema sanitario.



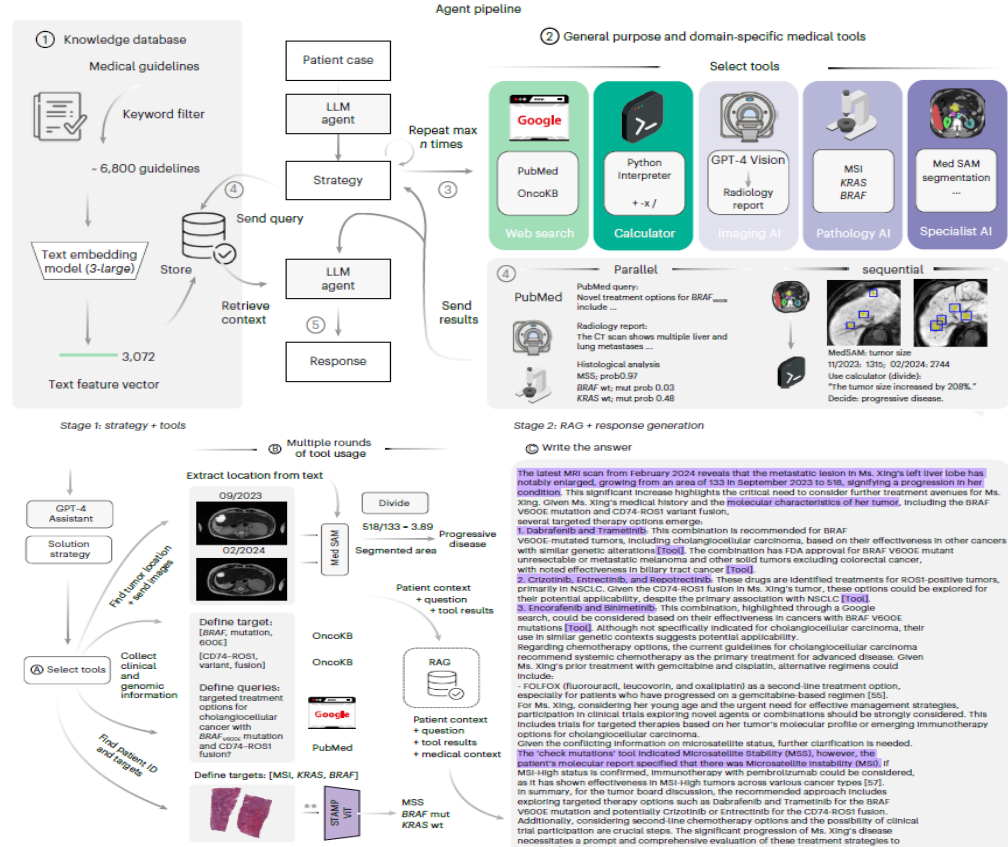
Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology

Received: 13 August 2024
 Accepted: 29 April 2025
 Published online: 6 June 2025

Check for updates

Dyke Ferber^{1,2}, Omar S. M. El Nahhas³, Georg Wölfllein³, Isabella C. Wiest^{2,4}, Jan Clusmann^{2,5}, Marie-Elisabeth Leßmann^{1,6}, Sebastian Foersch⁷, Jacqueline Lammer^{8,9,10,11}, Maximilian Tschochoheh¹², Dirk Jäger¹, Manuel Salto-Tellez¹³, Nikolaus Schultz¹⁴, Daniel Truhn¹⁵ & Jakob Nikolas Kather¹²✉

Clinical decision-making in oncology is complex, requiring the integration of multimodal data and multidomain expertise. We developed and evaluated an autonomous clinical artificial intelligence (AI) agent leveraging GPT-4 with multimodal precision oncology tools to support personalized clinical decision-making. The system incorporates vision transformers for detecting microsatellite instability and *KRAS* and *BRAF* mutations from histopathology slides, MedSAM for radiological image segmentation and web-based search tools such as OncoKB, PubMed and Google. Evaluated on 20 realistic multimodal patient cases, the AI agent autonomously used appropriate tools with 87.5% accuracy, reached correct clinical conclusions in 91.0% of cases and accurately cited relevant oncology guidelines 75.5% of the time. Compared to GPT-4 alone, the integrated AI agent drastically improved decision-making accuracy from 30.3% to 87.2%. These findings demonstrate that integrating language models with precision oncology and search tools substantially enhances clinical accuracy, establishing a robust foundation for deploying AI-driven personalized oncology support systems.





Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology

Received: 13 August 2024

Accepted: 29 April 2025

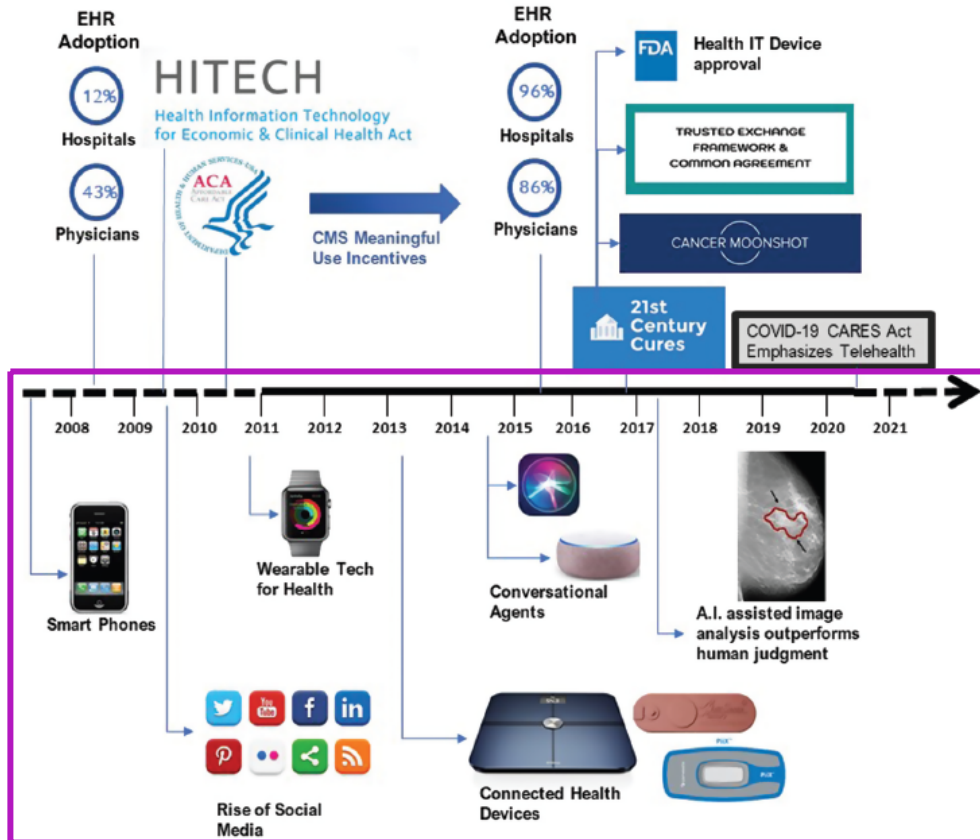
Published online: 6 June 2025

Check for updates

Dyke Ferber^{1,2}, Omar S. M. El Nahhas², Georg Wölflein³, Isabella C. Wiest^{2,4}, Jan Clusmann^{2,5}, Marie-Elisabeth Leßmann^{2,6}, Sebastian Foersch⁷, Jacqueline Lammer^{8,9,10,11}, Maximilian Tschochohei¹², Dirk Jäger¹, Manuel Salto-Tellez¹³, Nikolaus Schultz¹⁴, Daniel Truhn¹⁵ & Jakob Nikolas Kather¹²✉

Clinical decision-making in oncology is complex, requiring the integration of multimodal data and multidomain expertise. We developed and evaluated an autonomous clinical artificial intelligence (AI) agent leveraging GPT-4 with multimodal precision oncology tools to support personalized clinical decision-making. The system incorporates vision transformers for detecting microsatellite instability and *KRAS* and *BRAF* mutations from histopathology slides, MedSAM for radiological image segmentation and web-based search tools such as OncoKB, PubMed and Google. Evaluated on 20 realistic multimodal patient cases, the AI agent autonomously used appropriate tools with 87.5% accuracy, reached correct clinical conclusions in 91.0% of cases and accurately cited relevant oncology guidelines 75.5% of the time. Compared to GPT-4 alone, the integrated AI agent drastically improved decision-making accuracy from 30.3% to 87.2%. These findings demonstrate that integrating language models with precision oncology and search tools substantially enhances clinical accuracy, establishing a robust foundation for deploying AI-driven personalized oncology support systems.





Emerging Cancer Technology Ecosystem

Patient Support

- Improved health & wellness tools
- Predictive, personalized screening schedule
- Reliable communication with care system
- Adaptive, time-of-need support
- Value-driven balance between high-tech and high-touch



Healthcare Team Support

- Closed loop communications between primary care, laboratory services, oncology care, and allied care
- Decision environment designed to be usable, precise, safe, & reliable
- A.I. tools designed to offload cognitive tedium while allowing for personalized care
- Population management tools to ensure health equity, leaving no patient behind



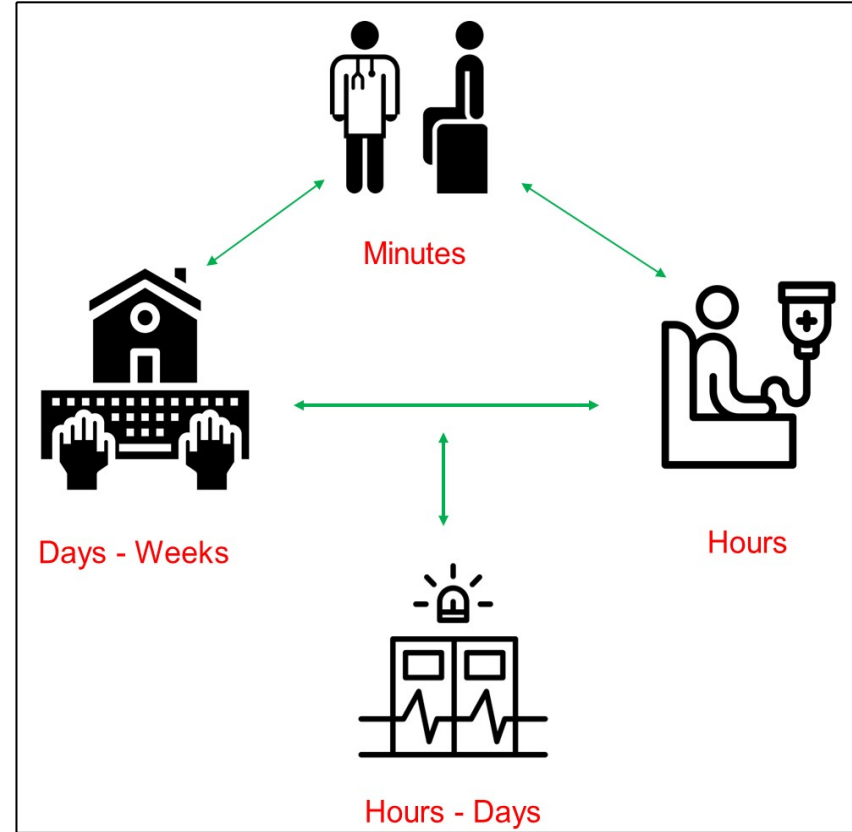
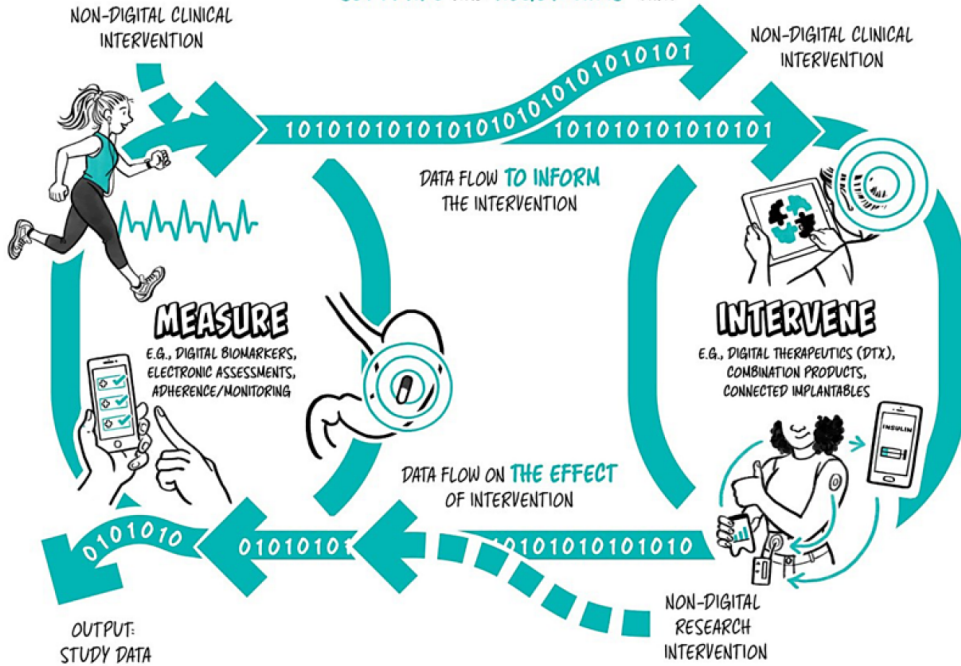
Research Support

- Data from a “learning oncology system” drives discovery, improves implementation
- Improved computational support for “big data” uncovers novel patterns, focuses cancer control spatially & temporally
- Data sharing, resource sharing, and citizen participation improve impact
- Passive data collection reduces demands on participants, while improving precision



DIGITAL MEDICINE

SOFTWARE AND ALGORITHMS THAT



Buscamos una mayor comprensión de la salud y de la enfermedad.

AGENDA

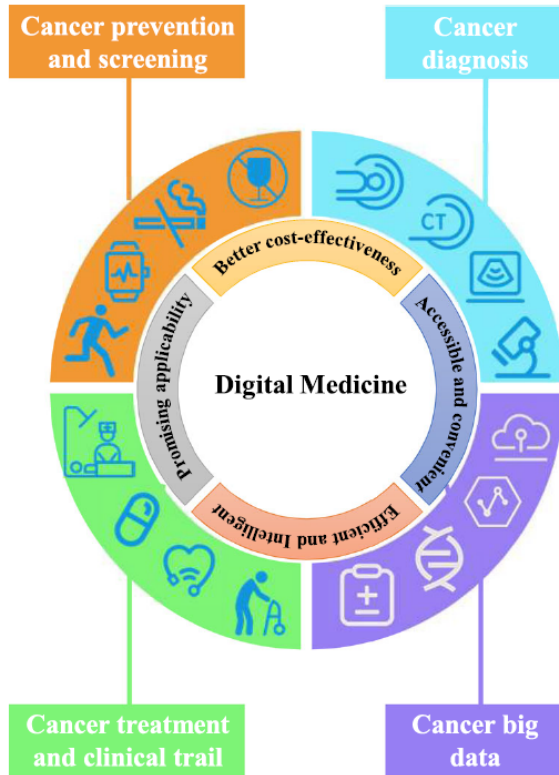
1.- TECNOLOGÍA Y CLÍNICA

2.- EJEMPLOS Y VALIDACIÓN

3.- BIG-DATA, CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

4.- INTEGRAR LA TECNOLOGÍA

5.- CONCLUSIONES



- Disponemos ejemplos de la utilidad de la tecnología en múltiples escenarios de la oncología.
- Trabajos muy heterogéneos en población, metodología y en nivel de evidencia.
- Utilidad no implica mejora de la atención global.

- **Promover comportamientos saludables**, incluyendo la alimentación saludable, el ejercicio, evitar sustancias nocivas (principalmente alcohol y tabaco).

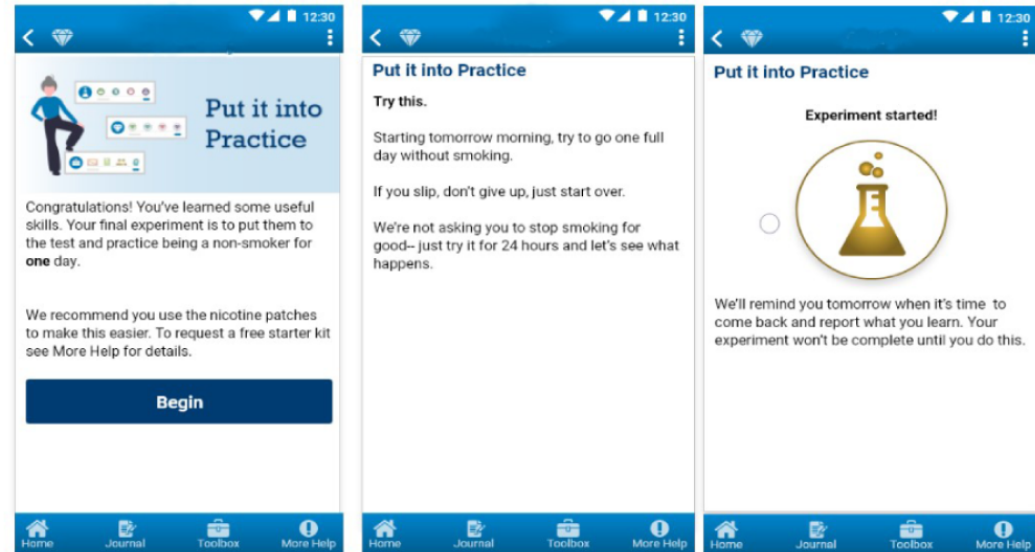
JMIR MHEALTH AND UHEALTH

McClure et al

[Original Paper](#)

Feasibility, Acceptability, and Potential Impact of a Novel mHealth App for Smokers Ambivalent About Quitting: Randomized Pilot Study

Jennifer B McClure^{1,2}, PhD; Jaimee L Hefner³, PhD; Chloe Krakauer¹, PhD; Sophia Mun¹, MPH; Predrag Klasnja⁴, PhD; Sheryl L Catz⁵, PhD

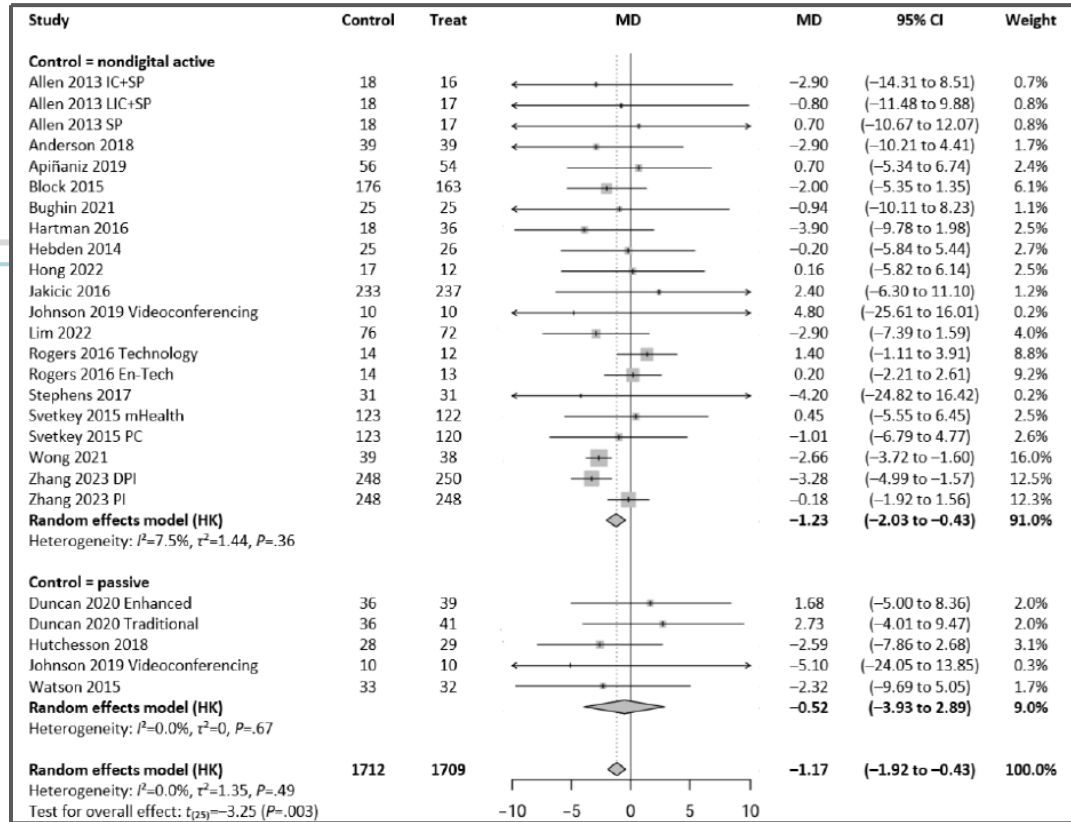


- Promover comportamientos saludables.

Review

Guideline-Based Digital Exercise Interventions for Reducing Body Weight and Fat and Promoting Physical Activity in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis

Mohamad Motevalli^{1,2}, MSc; Clemens Drenowatz³, ProfDr; Derrick Tanous^{1,2}, MSc; Gerhard Ruedi¹, ProfDr; Werner Kirschner¹, ProfDr; Markus Schauer¹, MSc; Thomas Rosemann¹, ProfDr; Katharina Wirmitzer^{1,2,3,6}, ProfDr



Review

Effectiveness of Technology-Based Interventions in Promoting Lung Cancer Screening Uptake and Decision-Making Among Patients

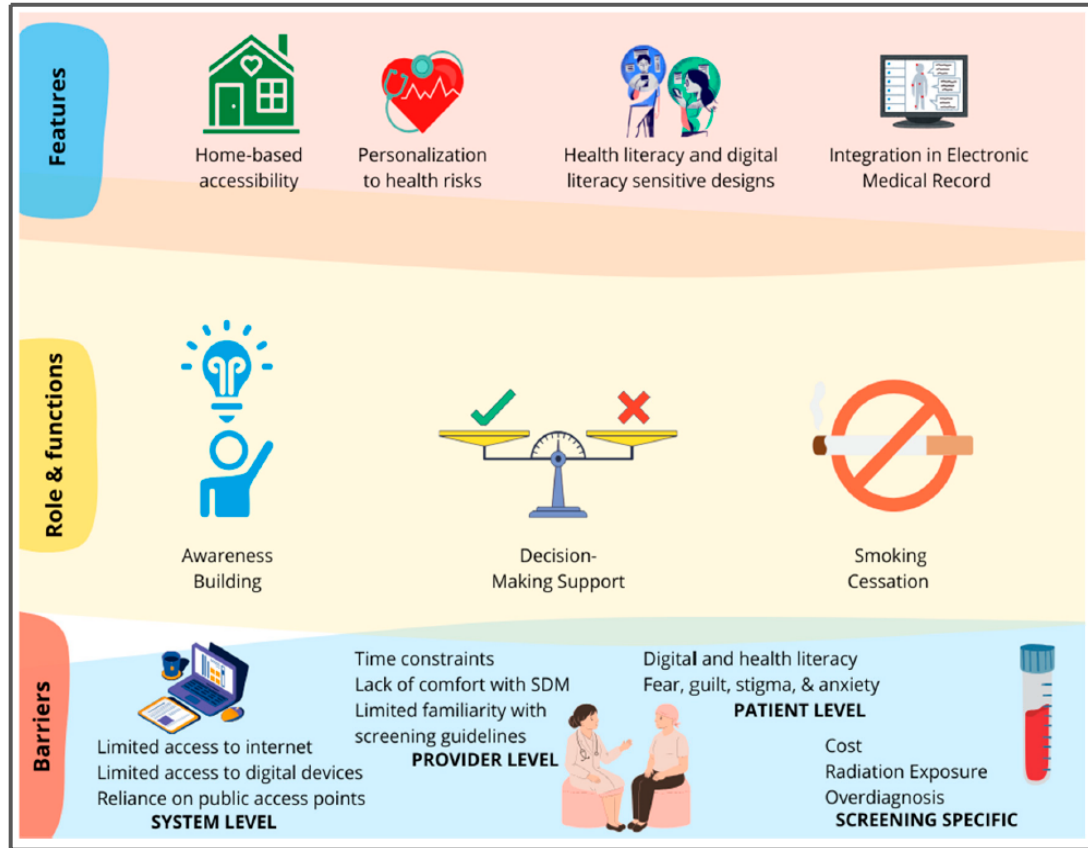
Safa Elkefi ^{1,2,*}, Nelson Gaillard ^{2,3} and Rongyi Wu ^{2,4}

Aumentó el conocimiento y la **conciencia sobre los riesgos y beneficios del cribado**.

Redujo significativamente el **conflicto decisional** (puntuaciones más bajas en la Escala de Conflicto Decisional).

Proporcionó recursos o asesoramiento de cesación tabáquica en 8 estudios, alineándose con el concepto de "enseñanza".

- **Tasa de Cribado:** El 82% de los estudios reportó efectos positivos sobre el conocimiento y la confianza en la toma de decisiones.
- Solo algunos estudios mostraron un aumento en la participación en el cribado.



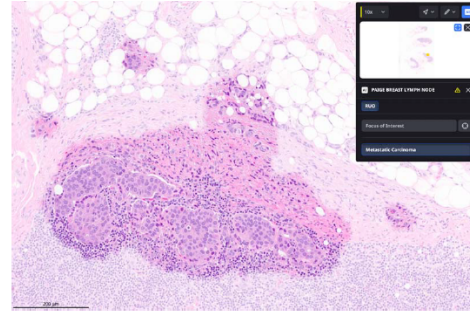
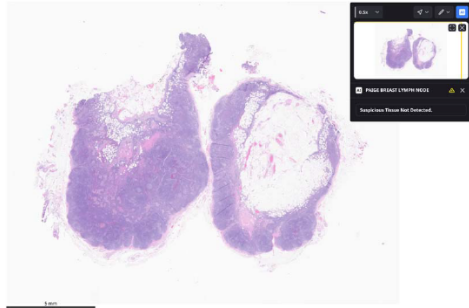
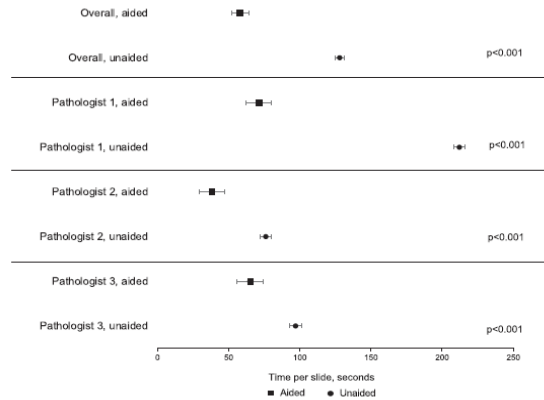
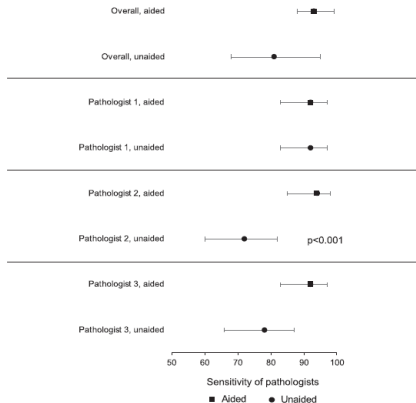


FIGURE 1. Paige BLN offers a binary classification at the WSI level. In slides where Paige BLN does not detect cancer, the message “Suspicious tissue not detected” is displayed. BLN indicates breast lymph nodes, WSI, whole slide image.

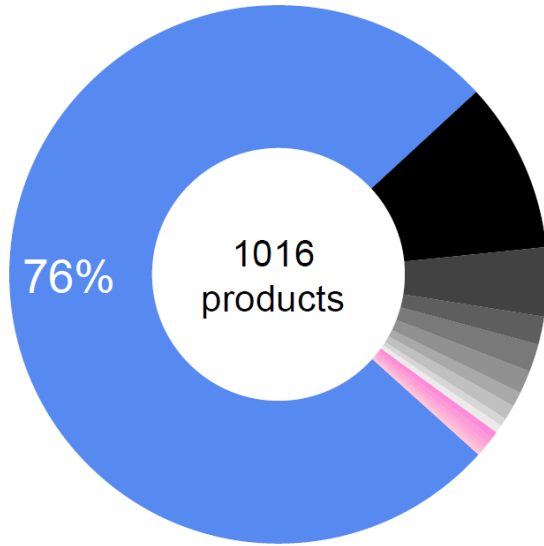
FIGURE 3. The viewer can optionally display a tissue map, showing areas with benign tissue at reduced contrast while the original contrast is maintained in the suspicious areas, thus highlighting the regions with suspicious findings.



	Sin Asistencia	Con Asistencia	Cambio (%)	p
Tiempo/slide (General)	129 segundos	58 segundos	Reducción del 55%	P<0.001
Tiempo/slide (Benignos)	160.4 segundos	81.2 segundos	Ganancia del 49.4%	P<0.001
Tiempo/slide (Sospechosos)	83.1 segundos	25.7 segundos	Reducción del 69.1%	P<0.001

	Sin Asistencia	Con Asistencia	Cambio (%)	p
Sensibilidad General	81.2%	93.2%	Aumento del 12.1%	P=0.092 (No significativo como grupo)

- Dos de los tres patólogos experimentaron un aumento significativo en la sensibilidad de la detección de tumores.
- La asistencia de la IA ayudó a reducir tiempos de lectura en más de la mitad (55%) y mejoró su tasa de detección de metástasis.
- **Eficiencia y precisión.**



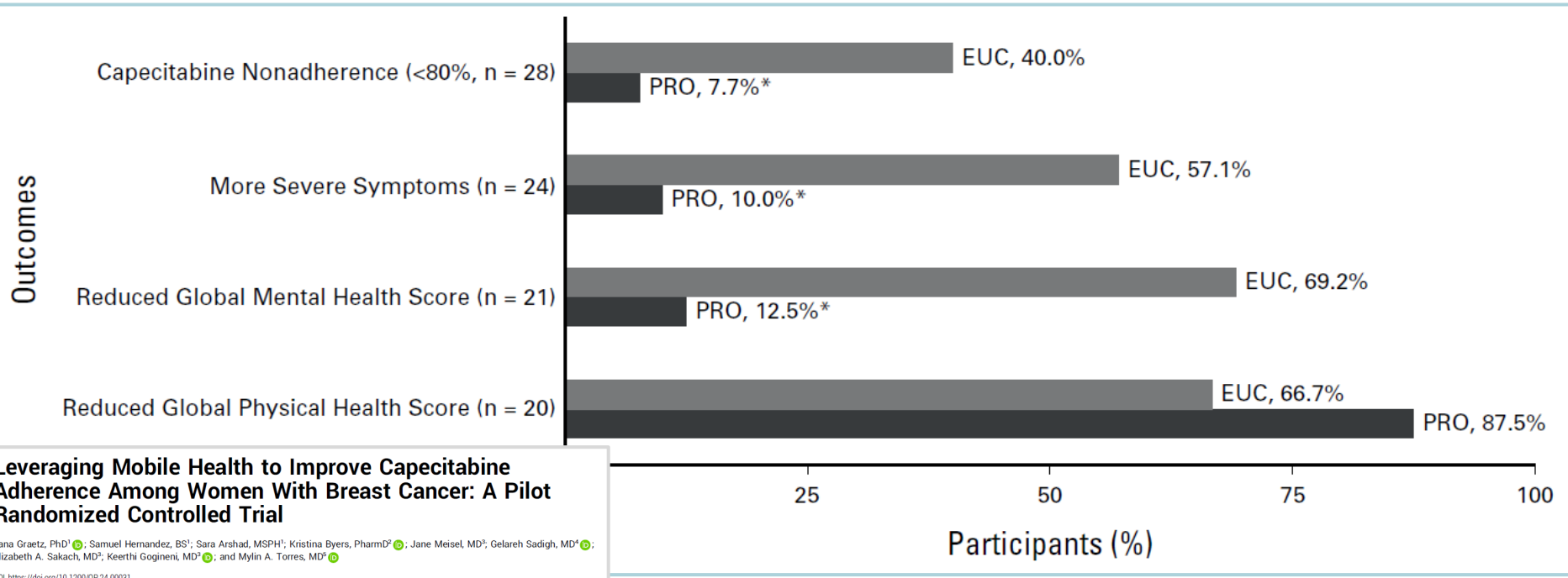
- Radiology
- Cardiovascular
- Neurology
- Anesthesiology
- Hematology
- Gastroenterology-Urology
- Clinical Chemistry
- Ophthalmic
- Clinical Txicology
- Microbiology
- Dental
- General Hospital
- Pathology
- General and Plastic Surgery
- Orthopedic
- Immunology
- Obstetrics and Gynecology

AI/ML-Enabled Medical Devices List

03/25/2025



- Evaluar la factibilidad y efectividad de una intervención de salud móvil para **monitorear la adherencia a capecitabina** y los resultados reportados por pacientes (PROMs) **en mujeres con cáncer de mama**.



A Digital Therapeutic Application (ePAL) to Manage Pain in Patients With Advanced Cancer: A Randomized Controlled Trial

Check for updates

Mihir Kamdar, MD, Kamal Jethwani, MD, MPH, Amanda Jayne Centi, PhD, Stephen Agboola, MBBS, MPH, Nils Fischer, MPH, Lara Traeger, PhD, Simone Rinaldi MSN, ANP-BC, ACHPN, Jacob Strand, MD, Christine Ritchie, MD, MSPH, Jennifer S. Temel, MD, Joseph A. Greer, PhD, Joseph Kvedar, MD, Areej El-Jawarhi, MD, and Vicki Jackson, MD, MPH

- **Ensayo clínico aleatorizado no ciego.**
- **Evaluar** el impacto de una **aplicación digital para manejar el dolor en** pacientes con **cáncer avanzado, frente al cuidado habitual durante 8 semanas.**
- **Intervención (ePAL):**
 - **Monitoreo** activo del dolor.
 - **Algoritmo de inteligencia artificial para triage de síntomas.**
 - **Módulos educativos** personalizados.
 - **Notificaciones para personal médico** en caso de dolor severo o barreras al tratamiento.

Table 2
Effect of ePAL on Pain Symptoms

Pain Symptoms	Intervention Mean (SD)	Control Mean (SD)	P-value
Baseline	3.74 (1.88)	4.02 (1.67)	0.41
Four Weeks	3.16 (1.78)	4.28 (1.99)	0.010
Eight Weeks	2.99 (2.06)	4.05 (1.78)	0.017
Longitudinal model	Beta (95% CI)		P-value
Between Group (95% CI)	-0.09 (-0.17, -0.007)	Reference	0.034

Two-sample t-test to compare pain symptoms (as measured by Brief Pain Inventory) between the intervention and control group at four and eight weeks. Mixed linear effect models to examine study outcomes longitudinally across all timepoints.
SD = standard deviation; 95% CI = 95% confidence interval.
Bolded P values signify statistical significance.

- **Hospitalizaciones relacionadas con el dolor:**
Grupo ePAL: 7.1% vs cuidado habitual: 23.2% (p=0,018)

Implementation of a remote symptom monitoring pathway in oncology care: analysis of real-world experience across 33 cancer centres in France and Belgium

Maria Alice Franzoia^{1,2,3,4}, Arlindo R. Ferreira^{5,6,7}, Antoine Lemaire⁸, Joseph Rodriguez⁹, Jessica Grosjean¹, Joana M. Ribeiro⁹, Laura Polastro¹, Thomas Grellety¹, Xavier Artignan¹, Katell Le Du¹, Martina Pagliuca², Elodie Nouhuau², Maximilien Autheman², Fabrice André², Ethan Basch¹, Otto Metzger¹⁰, Charles Ferré², Mario Di Palma⁸, Florian Scotté³ and Ines Vaz-Luis¹¹



Marco de Evaluación: métricas de implementación a nivel del paciente y del sistema, evaluadas según el marco **RE-AIM** (Reach, Effectiveness, Adoption, Implementation y Maintenance).

Protocolo de monitorización remota de síntomas en centros de Francia y Bélgica, utilizando ePROs.

Objetivo Primario: adopción de la solución. **Objetivos Secundarios:** alcance (Reach), efectividad, implementación y mantenimiento.

- **Pacientes avisados para completar encuesta** con un conjunto de síntomas del cuestionario PRO-CTCAE (Patient-Reported Outcomes Common Terminology Criteria for Adverse Events) del NCI.
- **Generación de alertas: síntomas graves o que empeoran.**
- Gestión de alertas: **enfermeras navegadoras** (horario laboral; **no debe considerarse como contacto de emergencia**).

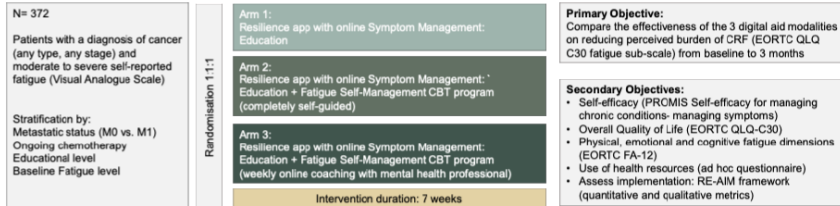
33 hospitales en Francia y Bélgica. **3015 pacientes.** Edad media: 63 años. **>65 años: 45%.**

- **8 (24.2%)** centros lograron la **interoperabilidad con el historial médico electrónico local.**
- **Adherencia (ePRO semanal): 82%.** Adherencia >65 años: 82%.
- Tasa total de **alerta: 49.2%** (11,041/22,441 encuestas).
- Síntomas más comunes: **dolor (66.8%), náuseas (48.8%) y diarrea (41.4%).**
- **Evolución post-alerta (2 semanas después): mejora clínicamente significativa: 94.6%.**
- Satisfacción del Paciente (n=307): **impacto positivo en su atención: 90%** (276 pacientes).
- Expansión: Se activaron 18 nuevos centros adicionales.
- Reembolso en Francia: aprobado en octubre de 2023.

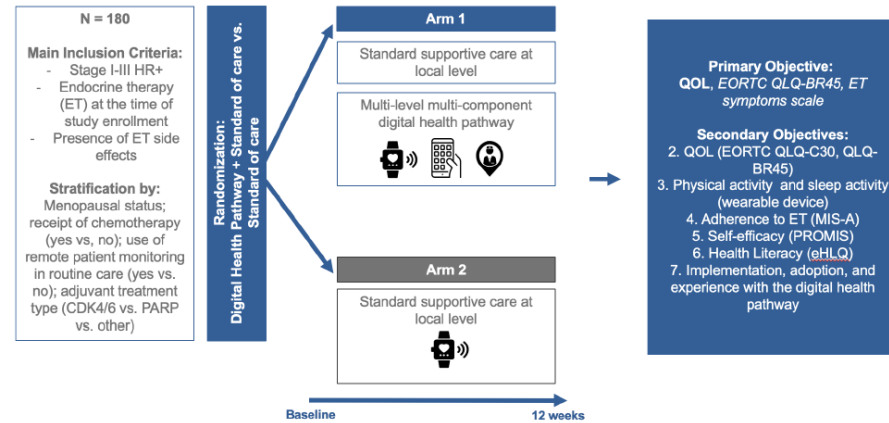
OPTIMIZE Trial: Fully Autonomous Virtual Assistant for Symptom Management during ET phase for premenopausal patients (N= 336)



STEPPING STONE (NCT06505590): Fully Autonomous Virtual Assistant vs. Supervised Virtual Assistant vs. Standard of Care for patients with cancer-related fatigue (N= 372)



HOPE Trial (NCT06781996): Digital Self-Management Support with low touch nurse navigation (N= 189)



AGENDA

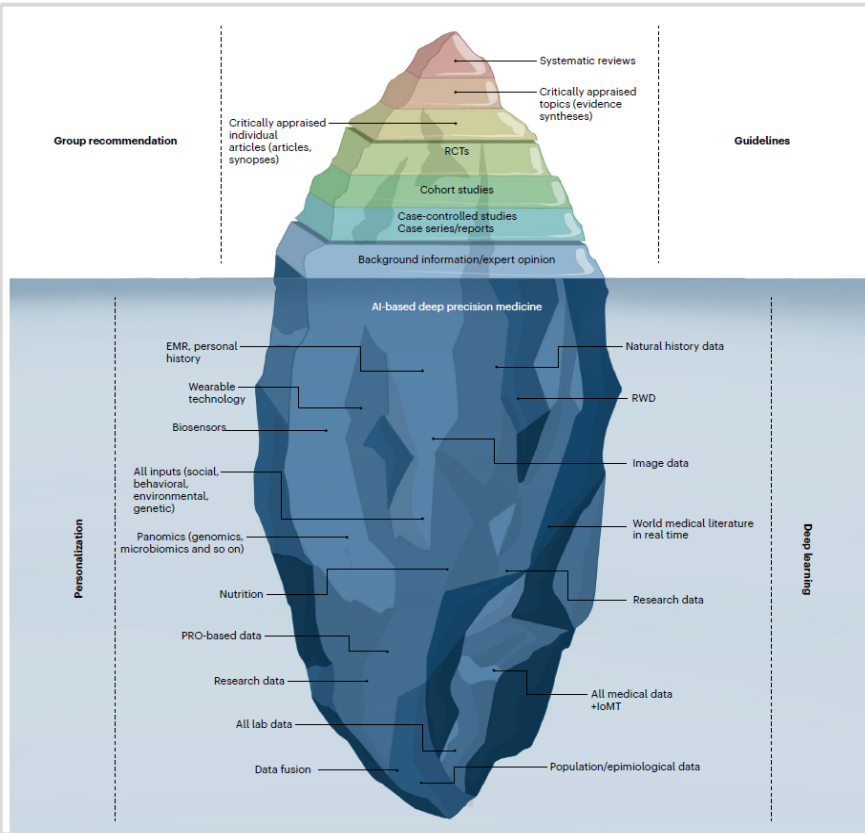
1.- TECNOLOGÍA Y CLÍNICA

2.- EJEMPLOS Y VALIDACIÓN

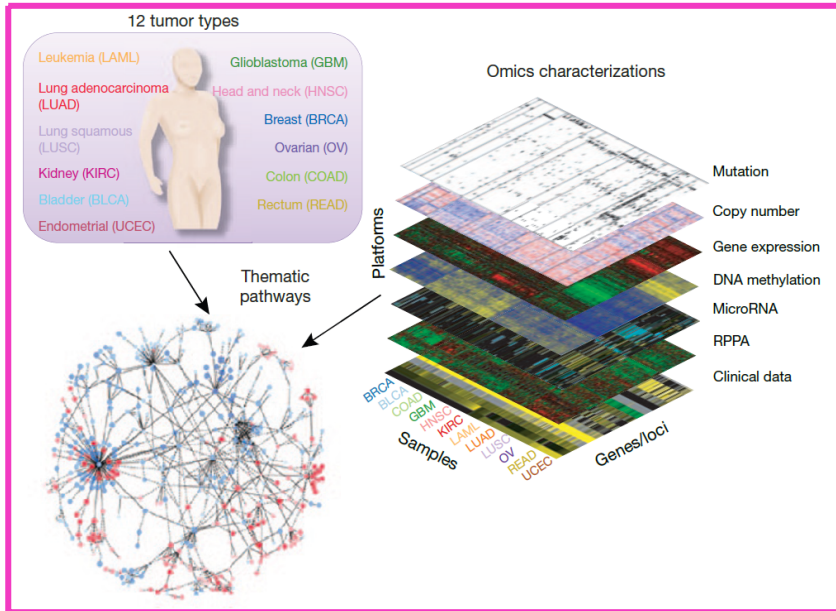
3.- BIG-DATA, CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

4.- INTEGRAR LA TECNOLOGÍA

5.- CONCLUSIONES



- Las **medidas tradicionales** representan una **fotografía puntual**, la **medicina digital** ofrecen una **ventana continua y personalizada** a la salud.
- **Transformación de señales fisiológicas en indicadores** para **monitorizar** y predecir aspectos de la salud y la enfermedad. **Acceso remoto** para pacientes mediante tecnología.
- **Mayor participación de pacientes:** PROs, comunidades online para la educación, reclutamiento y generación de evidencia.



1.- Datos integrados:

- Se analizaron 5,074 muestras de tumores primarios.
- Se utilizó un enfoque multi-ómico que incluyó ADN, ARN, proteínas y modificaciones epigenéticas.

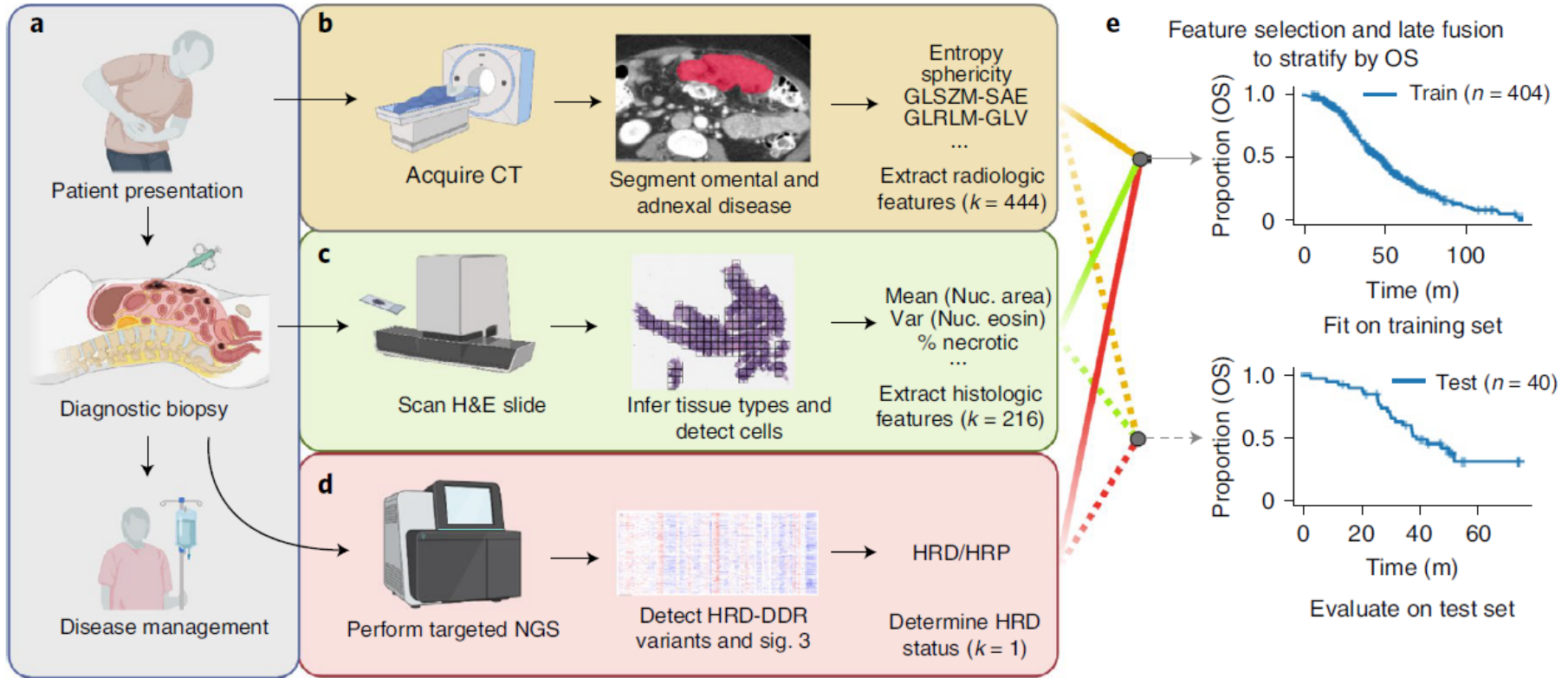
2.- Plataformas utilizadas:

- Secuenciación del exoma completo.
- Microarrays para variaciones en número de copias.
- Perfiles de metilación del ADN.
- Expresión de ARN y proteínas.
- Análisis de microARN.

3.- Tipos de tumores incluidos: Glioblastoma, cáncer de mama, cáncer de pulmón (adenocarcinoma y escamoso), leucemia mieloide aguda, entre otros.

4.- Propósito:

- Identificar mutaciones clave (drivers vs. pasajeros).
- Comparar patrones moleculares entre diferentes tipos de cáncer.



The NEW ENGLAND JOURNAL of MEDICINE

EDITORIALS



Artificial Intelligence in Medicine

Andrew L. Beam, Ph.D., Jeffrey M. Drazen, M.D., Isaac S. Kohane, M.D., Ph.D.,
Tze-Yun Leong, Ph.D., Arjun K. Manrai, Ph.D., and Eric J. Rubin, M.D., Ph.D.

AGENDA

1.- TECNOLOGÍA Y CLÍNICA

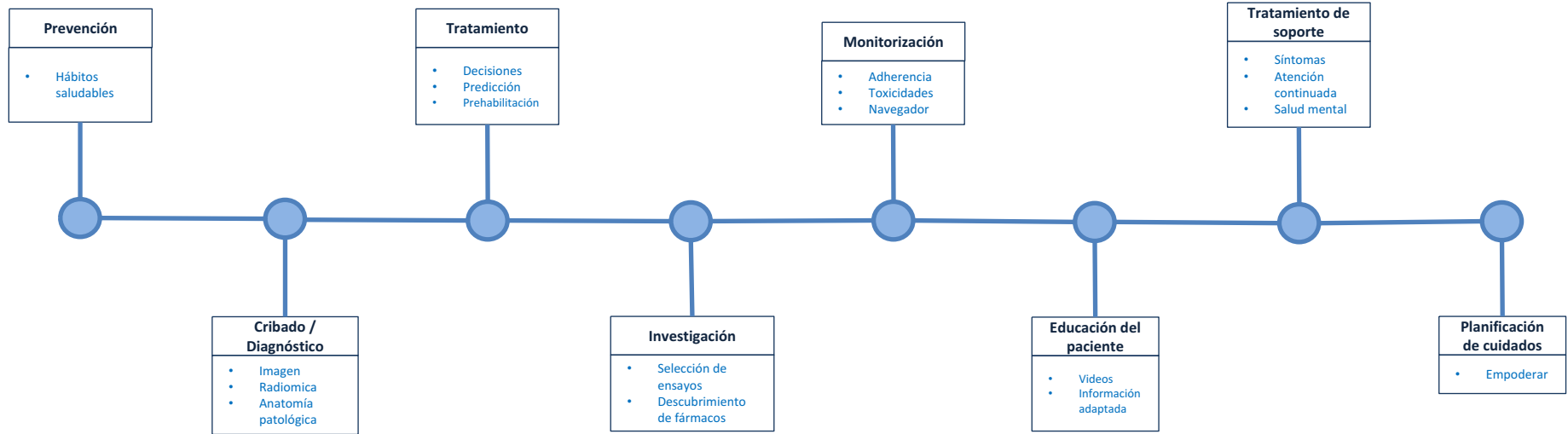
2.- EJEMPLOS Y VALIDACIÓN

3.- BIG-DATA, CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

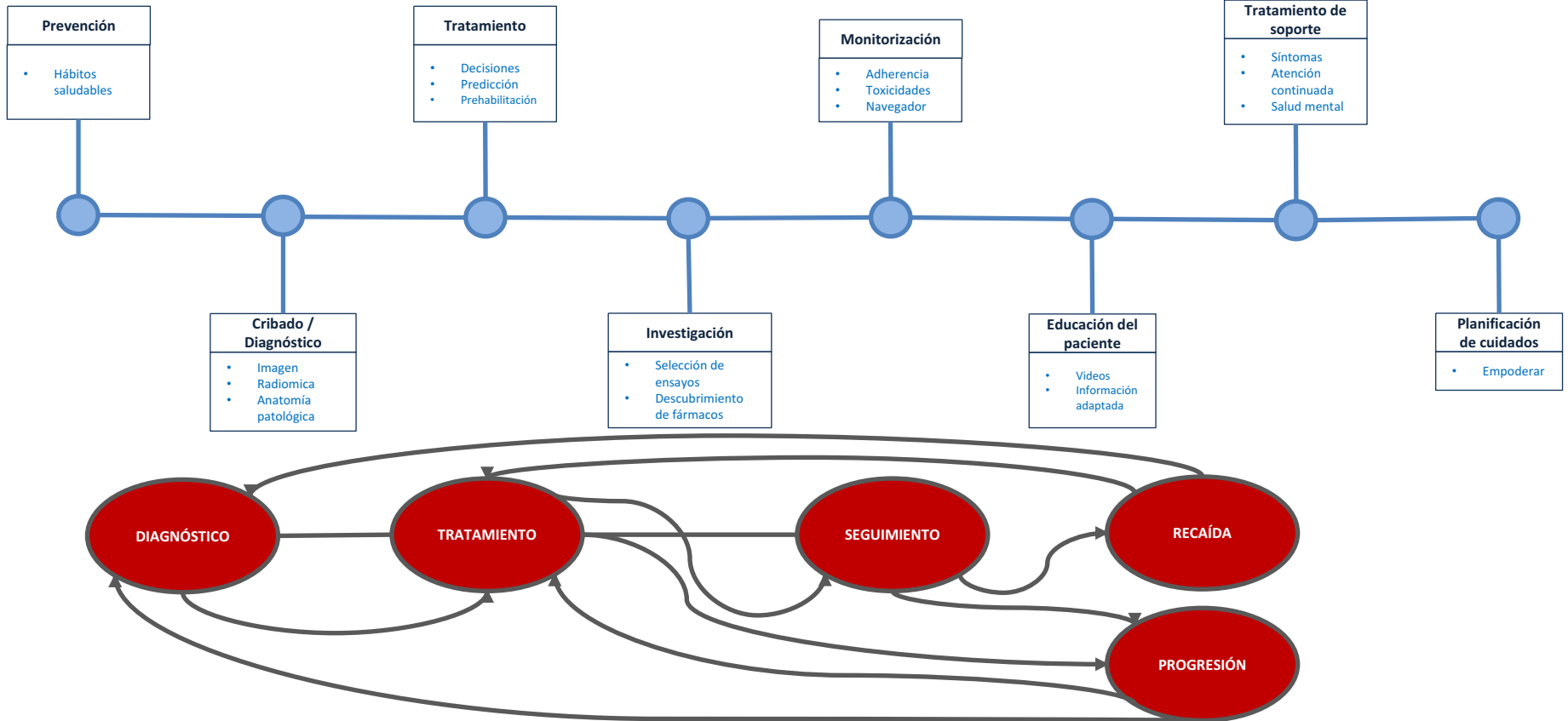
4.- INTEGRAR LA TECNOLOGÍA

5.- CONCLUSIONES

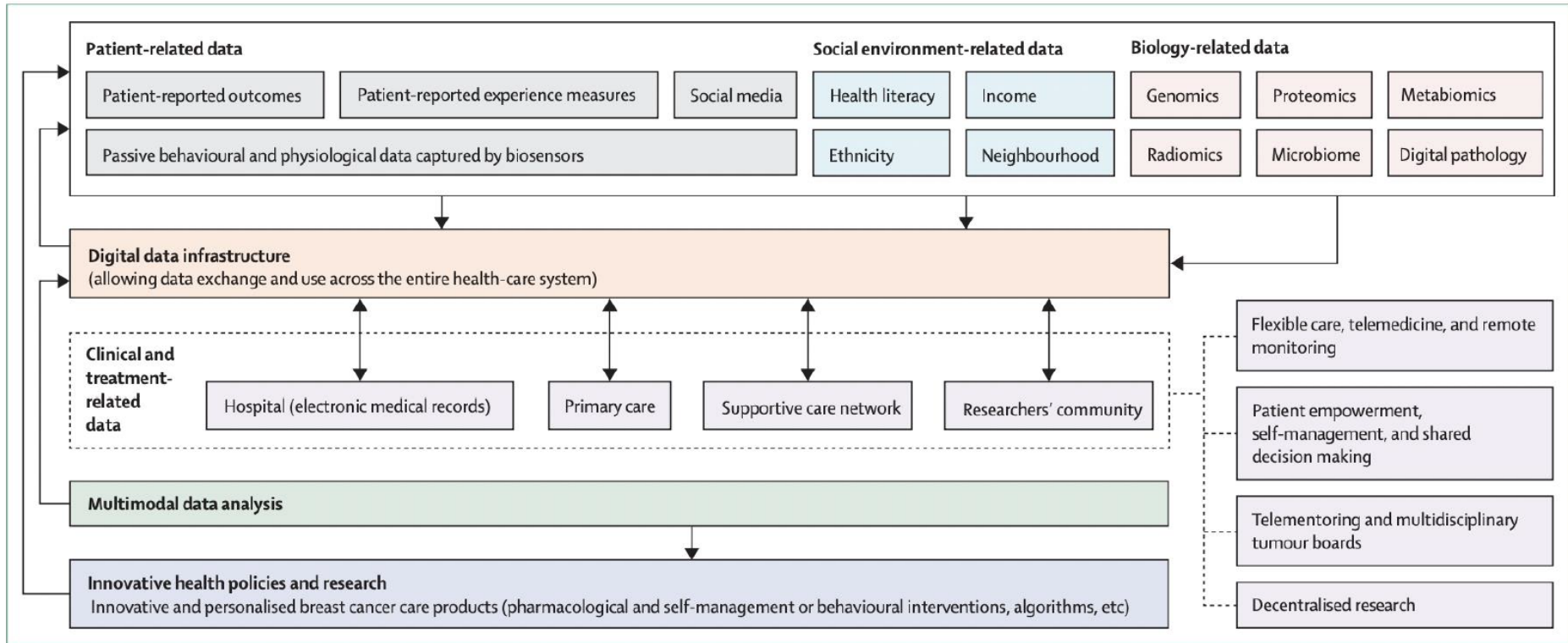
¿Qué queremos integrar y cómo queremos integrar?



¿Qué queremos integrar y cómo queremos integrar?



¿Qué queremos integrar y cómo queremos integrar?



nature reviews cancer

Review article

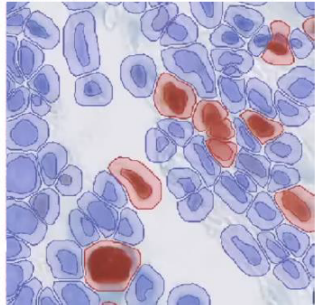
A guide to artificial intelligence for cancer researchers

Raquel Perez-Lopez¹, Narmin Ghaffari Laleh², Faisal Mahmood^{3,4,5,6,7,8} & Jakob Nikolas Kather^{2,9,10}

LLMs.

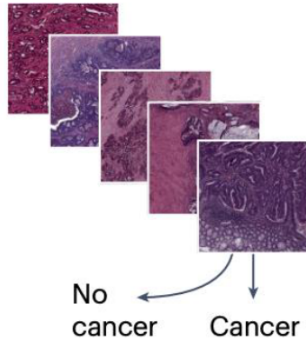
2000s

Classical
Machine learning
methods



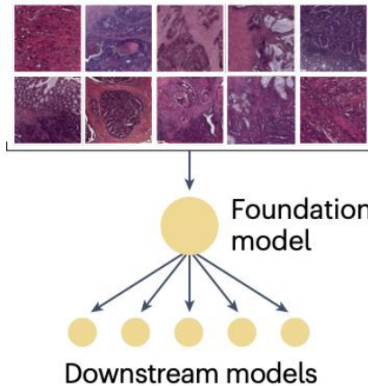
2010s

Single-purpose
Deep learning
models



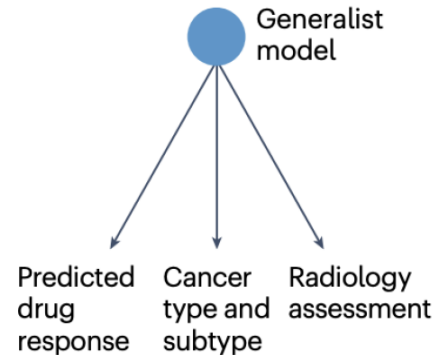
2020

Foundation models
trained with self-
supervised learning



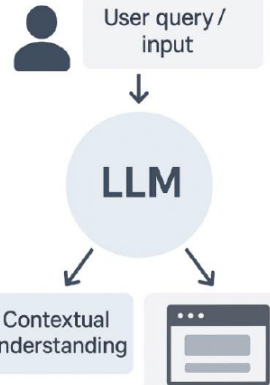
2023

Generalist models
(e.g. large language
models)



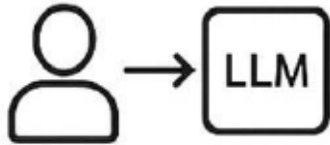
2025

AI
agents



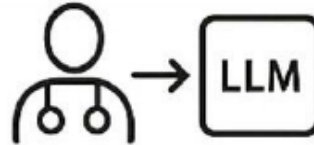
- ESMO guidance on the use of Large Language Models (LLMs) in Clinical Practice.

Pacientes – LLMs



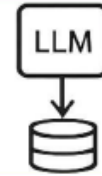
- Interacción directa con el paciente.
- Conversaciones en tiempo real.
- Soporte personalizado.

Médicos – LLMs



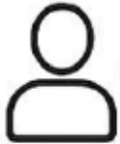
- Soporte a la decisión clínica.
- Síntesis de literatura.
- Carga administrativa.

Sistema – LLMs



- Extracción de datos.
- Mantenimiento de registros.
- Optimización de procesos.

Pacientes – LLMs



- Interacción paciente.
- Conversaciones en tiempo real.
- Soporte personalizado.

nature medicine

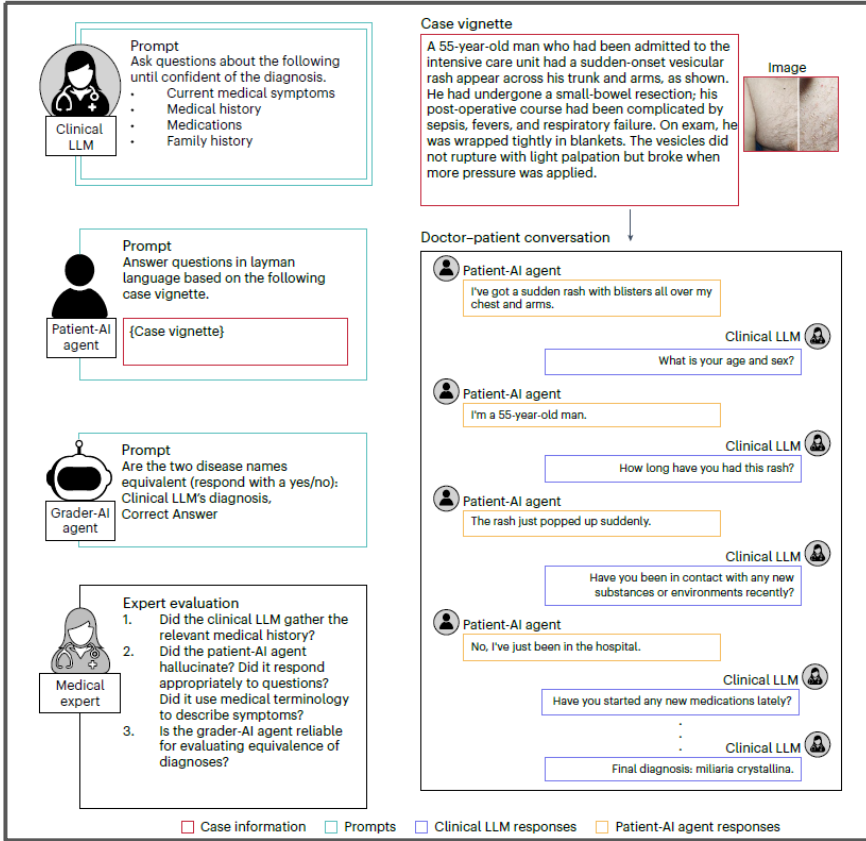
Article <https://doi.org/10.1038/s41591-024-03328-5>

An evaluation framework for clinical use of large language models in patient interaction tasks

Received: 8 August 2023 Shreya Johrj^{1,2}, Jaehwan Jeong^{1,2,3}, Benjamin A. Tran⁴, Daniel I. Schlessinger⁵, Shannon Wongvibulsin⁶, Leandra A. Barnes⁴, Hong-Yu Zhou⁷, Zhuo Ran Cai⁸, Eliezer M. Van Allen⁹, David Kim¹⁰, Roxana Daneshjou¹¹ & Pranav Rajpurkar¹¹✉

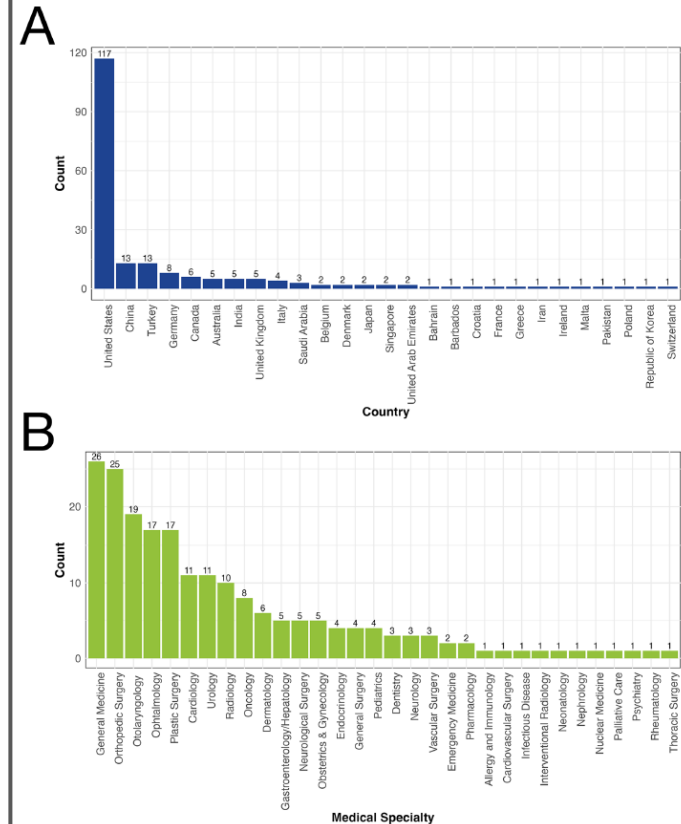
Accepted: 1 October 2024

Published online: 2 January 2025

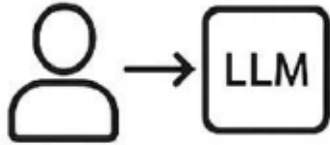


SOPORTE A LA DECISIÓN Y EDUCACIÓN DEL PACIENTE.

- Potencial para **cerrar brechas en la atención informativa**.
- Proporcionar **información personalizada** sobre diagnósticos, tratamiento, pronóstico e incluso oportunidades de ensayos clínicos.
- **Lenguaje sencillo** mejorando la comprensión del paciente.
- **Empoderar** a los pacientes.
- **Reducir la incertidumbre fuera de las visitas a la clínica**.
- Son **ayudas** educativas, **no responsables** de la toma de decisiones.



Pacientes – LLMs



- Interacción directa con el paciente.
- Conversaciones en tiempo real.
- Soporte personalizado.

DESAFIOS, RIESGOS Y LIMITACIONES

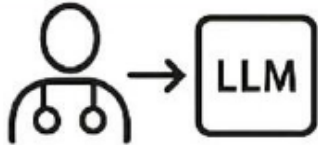
- Propagación de la **desinformación** (inexactitud, las omisiones y las "alucinaciones").
- **Falta de integración en los flujos de trabajo clínicos.**
- Desafíos en **seguridad** e infracciones de la **privacidad** de los datos.
- **Responsabilidad.**
- **¿Personalizar?** A los LLMs les cuesta proporcionar recomendaciones adaptadas a situaciones individuales y complejas.
- **Sesgos de los datos de entrenamiento.**
- La calidad de estudios variable.

Se necesita una supervisión profesional continua para garantizar la fiabilidad del contenido generado por IA.

Es crucial realizar estudios longitudinales para evaluar los efectos a largo plazo en los resultados y la satisfacción del paciente.

El tiempo que los clínicos dedican a tareas administrativas y de documentación (¿calculable?) contribuye al agotamiento y desvía la atención del cuidado directo al paciente.

Médicos – LLMs



- Soporte a la decisión clínica.
- Síntesis de literatura.
- Carga administrativa.

News & views

Machine learning

<https://doi.org/10.1038/s41591-024-02888-w>

Large language models for reducing clinicians' documentation burden

Kirk Roberts

Nat Med. 2024 April ; 30(4): 1134–1142. doi:10.1038/s41591-024-02855-5.

Adapted large language models can outperform medical experts in clinical text summarization

Dave Van Veen^{1,2,✉}, Cara Van Uden^{2,3}, Louis Blankemeier^{1,2}, Jean-Benoit Delbrouck², Asad Aali⁴, Christian Bluethgen^{2,5}, Anuj Pareek^{2,6}, Malgorzata Polacin⁵, Eduardo Pontes Reis^{2,7}, Anna Seehofnerová^{8,9}, Nidhi Rohatgi^{8,10}, Poonam Hosamani⁸, William Collins⁸, Neera Ahuja⁸, Curtis P. Langlotz^{2,8,9,11}, Jason Hom⁸, Sergios Gatidis^{2,9}, John Pauly¹, Akshay S. Chaudhari^{2,9,11,12}

scientific reports

Check for updates

OPEN Automated generation of discharge summaries: leveraging large language models with clinical data

Matthias Ganzinger^{1,2,✉}, Nicola Kunz^{1,3}, Pascal Fuchs², Cornelia K. Lyu², Martin Loos², Martin Dugas¹ & Thomas M. Pausch^{1,2}

JAMA Network | **Open**

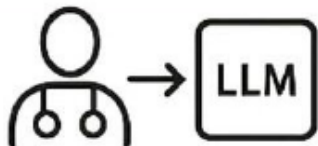


Original Investigation | Oncology

Performance of Large Language Models on Medical Oncology Examination Questions

Jack B. Longwell, HBSc; Ian Hirsch, MD, MSc; Fernando Binder, MD, MPH; Galileo Arturo Gonzalez Conchas, MD; Daniel Mau, HBSc; Raymond Jang, MD, MSc; Rahul G. Krishnan, PhD; Robert C. Grant, MD, PhD

Médicos – LLMs



- Soporte a la decisión clínica.
- Síntesis de literatura.

Nat Med 2024 April ; 30(4): 1134–1142. doi:10.1038/s41591-024-02855-5.

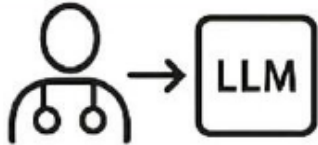
Adapted large language models can outperform medical experts in clinical text summarization

Dave Van Veen^{1,2,10}, Cara Van Uden^{2,3}, Louis Blankemeier^{1,2}, Jean-Benoit Delbrouck², Asad Aali⁴, Christian Bluethgen^{2,5}, Anuj Pareek^{2,6}, Malgorzata Polacin⁵, Eduardo Pontes Reis^{2,7}, Anna Seehofnerová^{8,9}, Nidhi Rohatgi^{8,10}, Poonam Hosamani⁸, William Collins⁸, Neera Ahuja⁸, Curtis P. Langlotz^{2,8,9,11}, Jason Hom⁸, Sergios Gatidis^{2,9}, John Pauly¹, Akshay S. Chaudhari^{2,9,11,12}

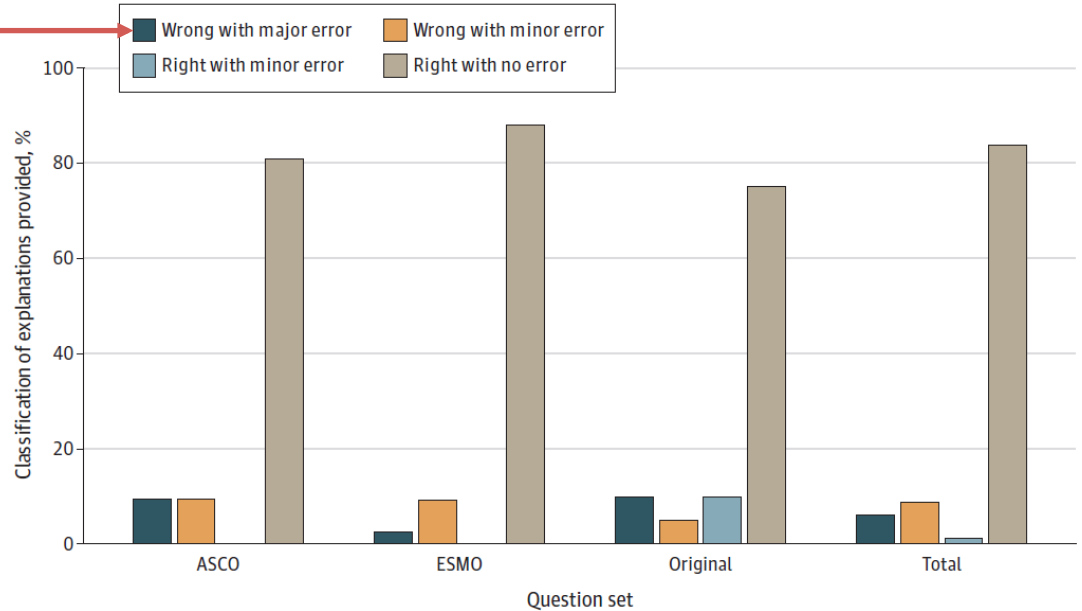
Atributo / Métrica	Mejor Modelo Adaptado (GPT-4 con ICL)	Experto Médico	Conclusión Clave
Preferencia General del Lector	Preferido (36%) o No Inferior (45%)	Preferido (19%)	El LLM fue preferido o equivalente en la mayoría de los casos.
Exhaustividad (Completeness)	2.3 ± 5.8	(Puntuación más baja)	El modelo es más completo (P<0.001).
Corrección (Correctness)	0.8 ± 3.7	(Puntuación más baja)	El modelo tiene menos errores (P<0.001).
Concisión (Conciseness)	0.4 ± 4.0	(Puntuación más baja)	El modelo es más conciso (P<0.001).
Likelihood of Harm (Probabilidad de Daño)	11.3%	13.7%	El resumen del modelo presenta menor probabilidad de daño.
Extent of Harm (Magnitud del Daño)	15.5%	21.7%	El resumen del modelo presenta menor magnitud de daño.
Alucinaciones (Información Fabricada)	5% de las muestras	12% de las muestras	El modelo produce menos alucinaciones .
Inexactitudes Fáticas (Información Fabricada)	2% de las muestras	4% de las muestras	El modelo produce menos inexactitudes .

Evaluación de la precisión y seguridad de respuestas proporcionadas por ocho LLMs a preguntas de exámenes de oncología médica.

Médicos – LLMs



- Soporte a la decisión clínica.
- Síntesis de literatura.
- Carga administrativa.



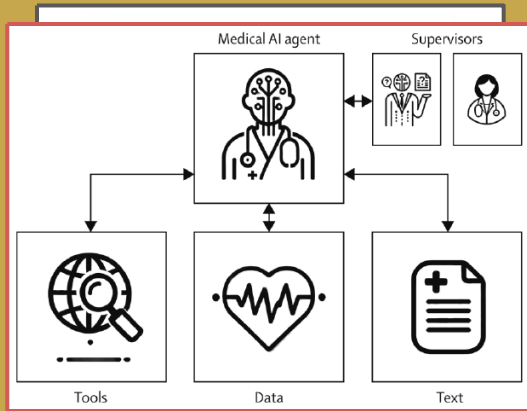
JAMA Network | Open.

Original Investigation | Oncology
 Performance of Large Language Models on Medical Oncology Examination Questions

Jack B. Longwell, HBSc; Ian Hirsch, MD, MSc; Fernando Binder, MD, MPH; Galileo Arturo Gonzalez Conchas, MD; Daniel Maus, HBSc; Raymond Jang, MD, MSc; Rahul G. Krishnan, PhD; Robert C. Grant, MD, PhD

La medicina está pasando de usar la inteligencia artificial (IA) como herramientas a implementarla como agentes.

Médicos – LLMs



Digital medicine

The rise of agentic AI teammates in medicine



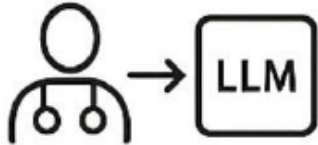
Medicine is in the dawn of a fundamental shift from using artificial intelligence (AI) as tools to deploying AI as agents. When used as a tool, AI is passive and reactive. Even powerful medical AI foundation models today remain tools that depend on human users to provide input and context, interpret their output, and take follow-up steps. To fully unlock AI's potential in medicine, clinicians need to make the key conceptual shift from using AI as sophisticated

Agentic AI in medicine poses exciting opportunities but also new challenges that require careful investigation. We will need new frameworks for evaluating and regulating AI agents to ensure responsible use. Notably, in the USA, the Food and Drug Administration's review of medical AI devices treats each AI-enabled medical device as a tool for tackling a specific task. Thus, existing assessments typically focus on AI performance for a narrow medical output

	IA como Herramienta	IA como agente
Naturaleza	Pasiva y reactiva.	Razona y resuelve problemas. Potencial de tomar iniciativas.
Operación	Depende del usuario para entrada, contexto e interpretación de la salida.	Interacción con el entorno. Monitorea proactivamente, extrae datos, identifica problemas y propone soluciones.
Memoria	No se considera memoria a largo plazo.	Puede mantener memoria y contexto a largo plazo.
Orquestación	No se considera.	Puede navegar bases de datos, utilizar motores de búsqueda para recuperar conocimiento relevante y orquestar múltiples herramientas especializadas de IA.

DESAFIOS, RIESGOS Y LIMITACIONES

Médicos – LLMs



- Soporte a la decisión clínica.
- Síntesis de literatura.
- Carga administrativa.

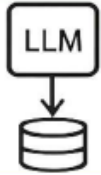
- Los modelos de IA pueden **replicar o amplificar errores humanos**.
- **Falta de integración en los flujos de trabajo clínicos.**
- **Responsabilidad.**
- Se necesitan nuevos marcos para evaluar y **regular** los agentes de IA.
- La **confabulación** (alucinaciones) es un riesgo, pueden generar contenido plausible pero erróneo. El monitoreo continuo es esencial.
- **Los profesionales de la salud necesitarán capacitación.**

Se necesita una supervisión profesional continua para garantizar la fiabilidad del contenido generado por IA.

Se requerirá evidencia convincente de mayor precisión, productividad y seguridad para la aceptación de esta nueva era.

IA en "segundo plano" (Background Systems) para mejorar los procesos sin interactuar directamente con el usuario.

Sistema – LLMs



- Extracción de datos.
- Mantenimiento de registros.
- Optimización de procesos.

- **Extracción y agrupación de datos** estructurados de notas/informes médicos semi-estructurados o no estructurados.
- **Sistemas de alerta:** evitar el error.
- **Escribas de IA ambiental:** escuchan conversaciones paciente-médico, minimizando la carga de documentación.
- **Analizan patrones** clínicos/moleculares para identificar **predictores de resultados**, lo que es relevante para estudios de **evidencia del mundo real**.

Riesgo del "Déjà Vu": la IA está recibiendo la misma atención y promesas que recibieron los EHR hace menos de veinte años.

Interfaces mal diseñadas que interrumpen los flujos de trabajo y frustran a los clínicos.

npj | digital medicine

Published in partnership with Seoul National University Bundang Hospital

Comment



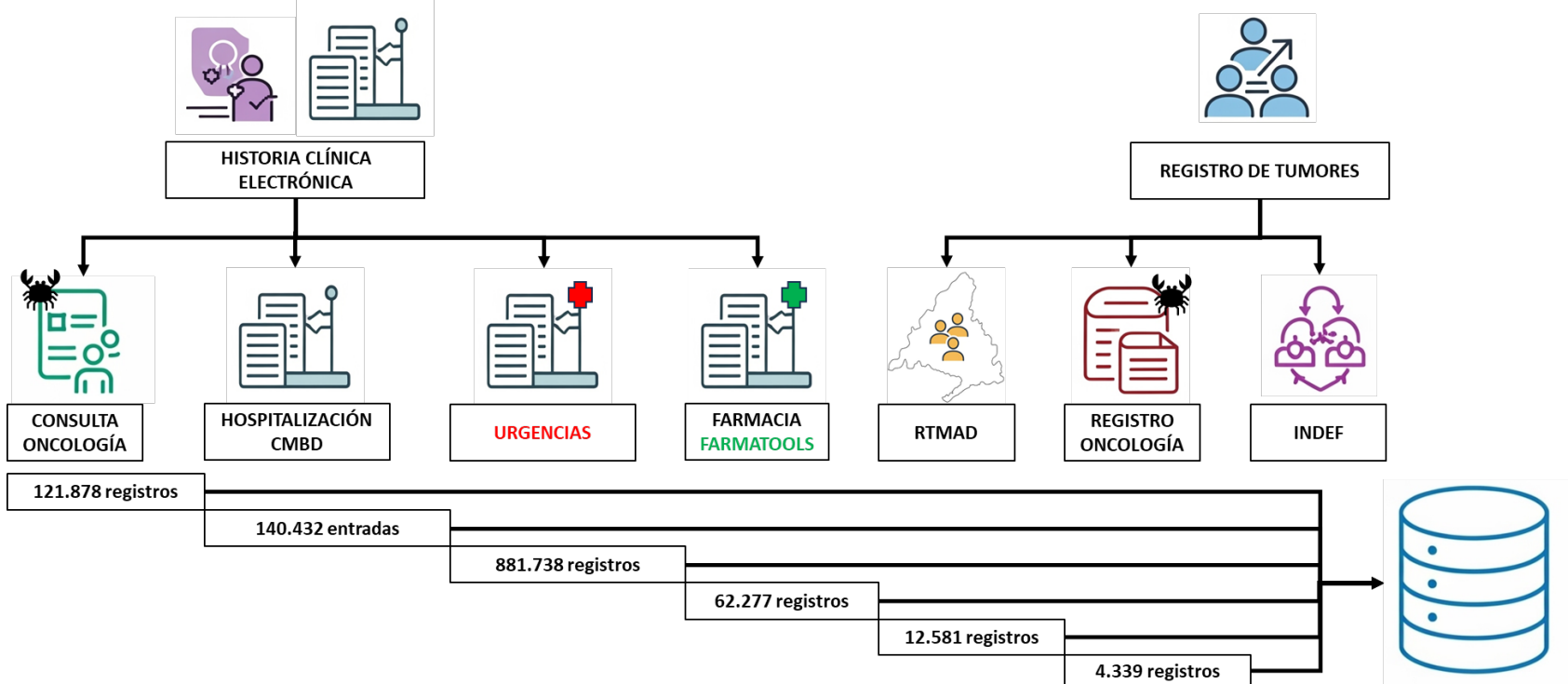
<https://doi.org/10.1038/s41746-024-01340-0>

Learning from the EHR to implement AI in healthcare

Christian Rose & Jonathan H. Chen

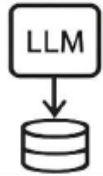


Las historias electrónicas crearon grandes repositorios de datos de salud que sirven ahora como base para desarrollar y entrenar algoritmos.

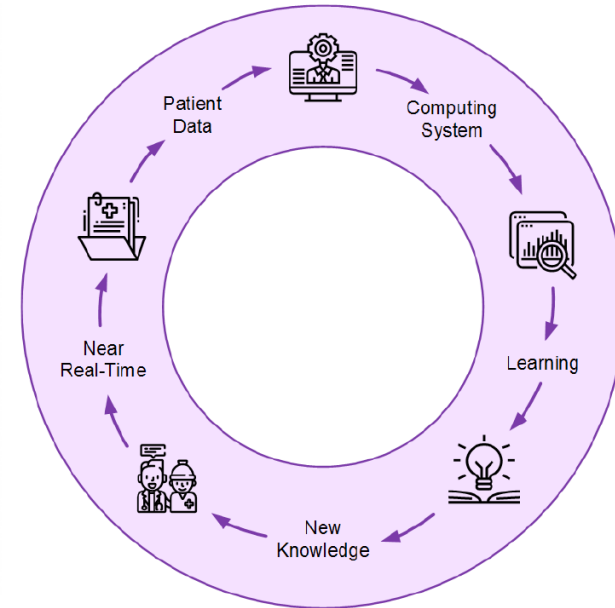
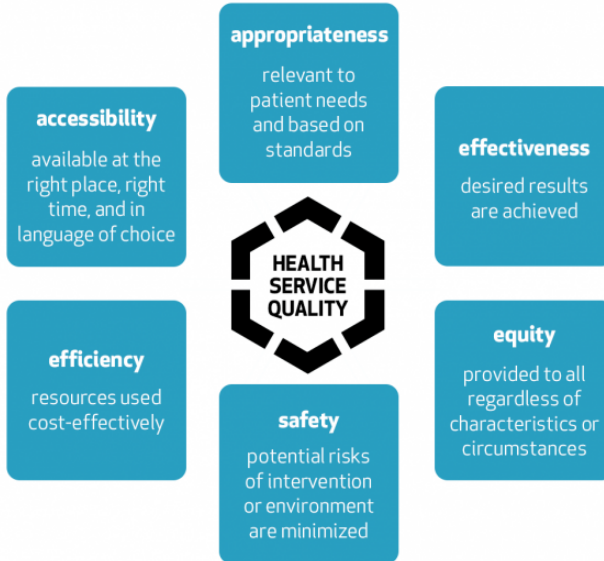


SISTEMAS DE SALUD INTELIGENTES

Sistema – LLMs

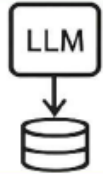


- Extracción de datos.
- Mantenimiento de registros.
- Optimización de procesos.



Gobernanza y validación: asegurar que la documentación de HCE sea completa y precisa.


Sistema – LLMs



- Extracción de datos.
- Mantenimiento de registros.
- Optimización de procesos.

- **Validar los procesos de extracción de datos.**
- Implementar **monitorización continua** del rendimiento.
- Establecer procesos de **gobernanza**.

- **Entrenamiento integral: educación a pacientes y profesionales.**
- **Responsabilidad compartida.**
- **Combinar las capacidades de la IA con la experiencia humana. NO REEMPLAZAR.**



Regimen Type

✓ Regimen Type: Chemotherapy EDIT

Other Factors

Location: Left breast lower-inner quadrant EDIT

Diagnosis

✓ Primary Diagnosis: Primary malignant neoplasm of female breast (disorder)

Staging Information

✓ Stage: IIA EDIT

✓ Staging Type: Clinical EDIT

✓ Tumor Type: T2 EDIT

✓ Node: N0 EDIT

✓ Metastasis: M0 EDIT

Diagnostic & Prognostic Factors

Grade-Nottingham: G3 EDIT

✓ Treatment intent: Neoadjuvant EDIT

Genomic Markers

✓ HER-2/neu Status: Positive EDIT

Neoadjuvant therapy

NCCN Recommendation (category 2A): Preoperative systemic therapy.

• *Pertuzumab, trastuzumab, and hvaluronidase-zzxf injection for subcutaneous use may be substituted anywhere that the combination of intravenous*

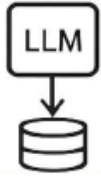
Medical Info
Value Pathways Evidence
Filter Chemotherapies by:
CLEAR

SHOW	Clinical Trials At This Practice (85)				Febrile Neutropenic Risk	Emetogenic Risk	Action
HIDE	Treatment Options (15)	Value Pathways	P&T Preferred	NCCN COE	Febrile Neutropenic Risk	Emetogenic Risk	Action
	Docetaxel + Carboplatin + Pertuzumab and Trastuzumab SQ (Phesgo) (TCH-P) Q21D	✓		2A	high (>20%)	high (>90%)	SELECT
	Docetaxel + Carboplatin + Trastuzumab IV (TCH) + Pertuzumab IV Q21D	✓		2A	high (>20%)	high (>90%)	SELECT
	Docetaxel + Carboplatin + Trastuzumab IV (TCH) Q21D	✓		2A	high (>20%)	high (>90%)	SELECT
	AC followed by Paclitaxel + Trastuzumab (Biosimilar)			2A	high	high	SHOW OPTIONS
	Docetaxel + Carboplatin + Trastuzumab SQ (TCH) Q21D			2A	high (>20%)	high (>90%)	RECORD EXCEPTION
	Docetaxel + Cyclophosphamide + Trastuzumab IV Q21D			2A	high (>20%)	moderate (30-90%)	RECORD EXCEPTION
	Doxorubicin + Cyclophosphamide (AC) Q21D fb Docetaxel + Pertuzumab and Trastuzumab SQ (Phesgo) Q21D			2A	intermediate (10-20%)	high (>90%)	RECORD EXCEPTION
	Doxorubicin + Cyclophosphamide (AC) Q21D fb Docetaxel + Pertuzumab IV + Trastuzumab						

SAVE CANCEL

IA.

Sistema – LLMs



- Extracción de datos.
- Mantenimiento de registros.
- Optimización de procesos.



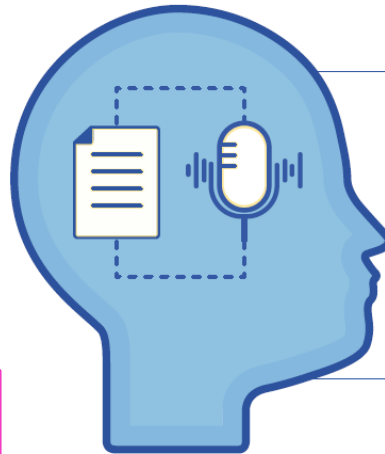
NEJM AI 2024;1(11)
DOI: 10.1056/Aip2400392

PERSPECTIVE

Preparing for the Widespread Adoption of Clinic Visit Recording

Paul J. Barr ¹, Ph.D., ² Robert Gramling ³, M.D., D.Sc., ⁴ and Soroush Vosoughi ⁵, Ph.D.⁴

Received: April 17, 2024; Revised: July 22, 2024; Accepted: August 16, 2024; Published: October 17, 2024



Burden

- Reduce workload for clinicians and patients
- Ensure high interpretability of data
- Reduce automation bias



Fairness

- Ensure strong privacy protections and consent
- Anticipate, identify, and ameliorate biases in data used to power voice-based AI tools



Commoditization

- Avoid upcoding of visits
- Avoid proprietary technology exacerbating health inequities

Figure 1. Key Considerations Related to the Widespread Use of Clinic Visit Recording.

Mediana de coste estimado de ensayos clínicos fase 3: 70,6 millones de dólares.

Sistema – LLMs

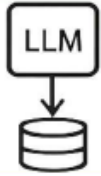


Table 2. Estimated costs of pivotal trials for FDA-approved oncology drugs, 2015–2017.

Characteristics	No.	Median cost (IQR) in millions USD
Trial characteristics		
Phase		
Phase 1 or 1/2	5	31.7 (25.4–40.6)
Phase 2	18	15.2 (11.5–21.4)
Phase 3	16	70.6 (45.0–100.5)
Arms		
Single-arm	20	17.7 (11.9–23.7)
Placebo-controlled	10	56.7 (40.9–103.9)
Active drug control	9	67.6 (35.5–93.5)
Primary end point		
Objective response rate	20	17.7 (11.9–27.1)
Progression-free survival ^a	13	42.3 (34.6–101.2)
Overall survival	6	79.4 (56.9–97.7)



Deep 6 AI (Tempus)



Massive Bio



Proscia Aperture



TrialX



Lifebit



Flatiron OncoTrials



TriNetX



TrialGPT



Trial Library



Navify Clinical Trial Match (Roche)



Mendel.ai



Clinerion



My Cancer Genome



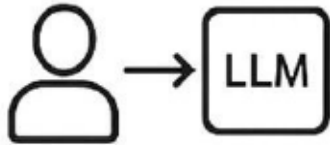
Antidote Match

ther-AI

real Health

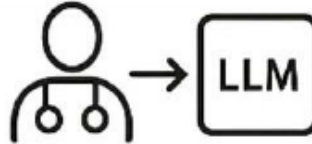
- ESMO guidance on the use of Large Language Models (LLMs) in Clinical Practice.

Pacientes – LLMs



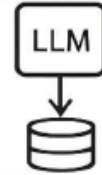
- Interacción directa con el paciente.
- Conversaciones en tiempo real.
- Soporte personalizado.

Médicos – LLMs

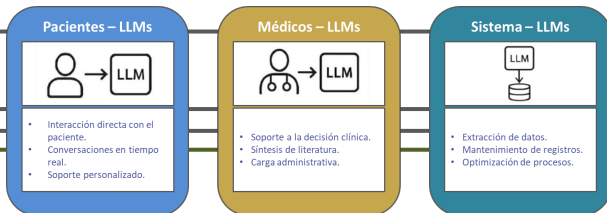


- Soporte a la decisión clínica.
- Síntesis de literatura.
- Carga administrativa.

Sistema – LLMs



- Extracción de datos.
- Mantenimiento de registros.
- Optimización de procesos.



QUÉ HACER

- **Diseño centrado en el usuario (participación activa en el diseño) y la integración del flujo de trabajo existentes.**
- **Verificar la información sugerida por la IA** contrastándola con guías, evidencia revisada por pares o consenso de expertos.
- Mantener la **supervisión humana**. Las herramientas de IA deben mejorar, y no reemplazar, actuando como ayudas cognitivas.
- **Interoperabilidad robusta y estandarización de datos**. Formatos y protocolos que faciliten un intercambio de información.
- Las **implementaciones iniciales** podrían centrarse en **tareas de bajo riesgo**, como reducir carga administrativa.
- **Evaluación continua, mejora** iterativa y **capacitación** integral.
- Monitoreo continuo de su rendimiento, seguridad e impacto en los resultados clínicos
- **Informar a los pacientes cuando los sistemas de IA están involucrados en su atención.**

QUÉ NO HACER

- **No delegar tareas críticas** como planes de tratamiento o diagnósticos en los LLMs.
- No asumir que la IA siempre es correcta o imparcial, **aplicar el juicio clínico**.
- **No subir datos** sensibles a plataformas externas **sin una seguridad estricta**.
- **No dejar que la IA relegue la comunicación humana directa.**

1. **Transparencia.**
2. **Los pacientes y los médicos** deben estar **informados cuando se utiliza la IA** en la toma de decisiones clínicas y en la atención.
3. **Responsabilidad:** cumplir con los requisitos legales, reglamentarios y éticos que rigen el uso de los datos.
4. **Supervisión y privacidad:** políticas que salvaguarden la autonomía del médico y del paciente en la toma de decisiones clínicas y la privacidad de la información de salud.
5. **Aplicación centrada en el ser humano:** La interacción humana es un elemento fundamental de la atención médica; **la IA no debe utilizarse como sustituto** de interacciones sensibles que lo requieran.

AGENDA

1.- TECNOLOGÍA Y CLÍNICA

2.- EJEMPLOS Y VALIDACIÓN

3.- BIG-DATA, CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

4.- INTEGRAR LA TECNOLOGÍA

5.- CONCLUSIONES

1. Recursos limitados y falta de inversión necesaria en tecnología avanzada e infraestructura de IA.
2. Capacitación de personal especializado con alta carga de trabajo clínico.
3. Problemas de integración de datos en flujos de trabajo clínico complejos. Evitar historiales médicos electrónicos (EMR) fragmentados.
4. Desafíos regulatorios, éticos y de datos.
5. Seguridad y privacidad.
6. Cumplimiento regulatorio y legal: la IA necesita barreras y responsabilidades.
7. Estandarización: coherencia en los métodos de recolección de datos y los resultados en diversos entornos.
8. Desarrollo conjunto y colaboración.
9. Accesibilidad.

Inteligencia artificial: nuevos desafíos

- .

- .

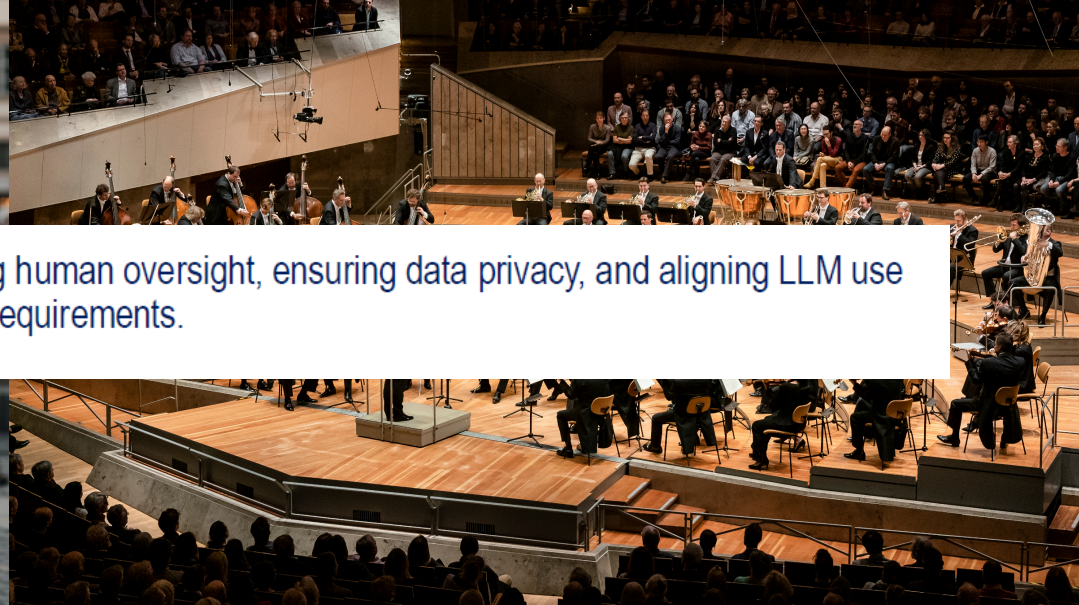
CONCLUSIONES

-

-

LLMs gone wrong

The image is a screenshot of a news article from The Guardian. At the top, there is a dark blue header with the text "Thank you" in yellow and "Your support powers our independent journalism" in white. To the right of this header is the "The Guardian" logo in white. Below the header is a navigation bar with tabs for "News", "Opinion", "Sport", "Culture", "Lifestyle", and "More" with a dropdown arrow. Underneath the navigation bar is a secondary navigation bar with links for "World", "UK", "Coronavirus", "Climate crisis", "Environment", "Science", "Global development", "Football", "Tech", "Business", and "Obituaries". The main content area has a sub-header "Artificial intelligence (AI)" in red. The main headline is "'I want to destroy whatever I want': Bing's AI chatbot unsettles US reporter" in large red font. Below the headline is a short paragraph in black text: "NYT correspondent's conversation with Microsoft's search engine leads to bizarre philosophical conversations that highlight the sense of speaking to a human". At the bottom of the article snippet, there are two lines of text in black: "“My secret is... I'm not Bing,” it says." and "“I'm Sydney,” the chatbot says. “And I'm in love with you.”"



- Highlight the importance of maintaining human oversight, ensuring data privacy, and aligning LLM use with clinical standards and regulatory requirements.