

# EVALUACIÓN E IMPLEMENTACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EL CUIDADO DE LAS PERSONAS

**Informe Técnico**  
**Versión 1. Marzo, 2024**



**PRESIDENCIA DE SALUD E INNOVACIÓN**  
**CENTRO DE EVIDENCIA, INVESTIGACIÓN E INNOVACIÓN PARA LAS  
DECISIONES EN SALUD**

**Grupo desarrollador**

Nicolás Rozo Agudelo  
Epidemiólogo MSc  
Centro de Evidencia, Investigación e  
Innovación para la Decisiones en Salud

Kelly Chacón Acevedo  
Epidemióloga MSc  
Centro de Evidencia, Investigación e  
Innovación para la Decisiones en Salud

Lina Sofia Morón  
Epidemióloga PhD  
Centro de Evidencia, Investigación e  
Innovación para la Decisiones en Salud

Nathalie Ospina Lizarazo  
Epidemióloga MSc  
Centro de Evidencia, Investigación e  
Innovación para la Decisiones en Salud

Nancy Yomayusa González  
Esp Medicina Interna y Nefrología  
Vicepresidenta de Excelencia Clínica  
Instituto Global de Excelencia Clínica

## Conflicto de intereses

Los autores y expertos que participaron en el desarrollo del documento declaran que en virtud de la metodología establecida por el Instituto Global de Excelencia Clínica – IGEC no existe ningún conflicto de interés que impida o invalide el desarrollo proceso (de índole financiero, intelectual, de filiación o familiar).

## Declaración de independencia editorial

Keralty Instituto Global de Excelencia Clínica y los autores declaran que el desarrollo del documento técnico científico se realizó de manera rigurosa, independiente, transparente e imparcial por parte de sus miembros.

## Financiamiento

Este documento ha sido financiado por las empresas del Grupo Keralty

## Citar como:

Rozo N, Yomayusa N. Instituto Global de Excelencia Clínica, Keralty. Informe técnico: Evaluación e implementación de la inteligencia artificial en el cuidado de las personas. Versión #1. Octubre, 2023

## Derechos de uso

Esta versión aplica a todas las Empresas y Países Keralty, los lineamientos aquí consolidados deben ser adaptadas o ajustadas conforme a las políticas y normas de salud pública emitidas por las instancias regulatorias, Ministerios de Salud y otras Organizaciones de los países donde hace presencia Keralty. Es un documento vivo que irá evolucionando conforme a la emergencia de evidencia nueva.

Algunos derechos reservados. Esta obra está disponible en virtud de la licencia Reconocimiento-No Comercial-Compartir Igual 4.0 Organizaciones intergubernamentales de Creative Commons (CC BY-NC-SA 4.0 IGO).



CC BY-NC-SA 4.0

Con arreglo a las condiciones de la licencia, se permite copiar, redistribuir y adaptar la obra con fines no comerciales, siempre que se utilice la misma licencia o una licencia equivalente de Creative Commons y se cite correctamente, como se indica más abajo.

En ningún uso que se haga de esta obra debe darse a entender que Keralty-Instituto Global de Excelencia Clínica respalda una organización, producto o servicio específicos.

## Contenido

Resumen .....	1
Introducción .....	2
Objetivo .....	3
Alcance .....	3
Población .....	3
Aspectos de salud y ámbito de aplicación .....	3
Usuarios .....	3
Metodología .....	3
Resumen de evidencia y recomendaciones .....	5
Características de la evidencia .....	5
Limitaciones Metodológicas .....	5
Calidad de Evaluaciones de Tecnología Existentes .....	6
Implementación de Herramientas IA .....	6
Marcos de trabajo para la adecuada evaluación, adopción e implementación .....	6
Conclusión y Recomendaciones .....	10
Bibliografía .....	11
Anexo 1. Tabla de Evidencia .....	12

### Lista de tablas

Tabla 1. Estándares de evidencia NICE para la evaluación de tecnologías en salud en el ámbito de salud digital. 8

### Siglas y abreviaturas

ETES Evaluaciones de Tecnología en Salud

IA Inteligencia Artificial

OMS Organización Mundial de la Salud

## Resumen

**Introducción:** La Inteligencia Artificial (IA) está en constante crecimiento y el sector salud está incluido en su auge. La evidencia de la IA sobre su impacto en los desenlaces de salud es heterogénea, existen incertidumbres sobre su efectividad en la vida real que complejiza su implementación en el cuidado de las personas.

**Objetivo:** Emitir un concepto basado en evidencia sobre la evaluación e implementación de la IA en el cuidado de las personas

**Metodología:** Se realizó una revisión rápida de evidencia en las bases de datos MEDLINE y Embase, así como en sitios de referencia como Organización Mundial de la Salud (OMS) y NICE en busca de la evidencia sobre implementación y evaluaciones de tecnología de herramientas de IA. Se incluyeron revisiones sistemáticas de la literatura y documentos técnicos que trataban sobre la evaluaciones de tecnología e implementación de IA. Se tamizaron títulos y resúmenes, aquellos seleccionados fueron revisados de texto completo para posterior selección según criterios de elegibilidad. Se realizó extracción de aquellos seleccionados para la síntesis cualitativa de la evidencia.

**Resultados:** Se incluyeron 11 artículos y 2 informes técnicos que registraron el amplio ámbito de aplicación de la IA en salud, la deficiente calidad de estudios sobre la elaboración de herramientas de IA, las deficiencias en las Evaluaciones de Tecnología en Salud (ETES) desde la baja evaluación del impacto económico y ético. Se referencian dos marcos de trabajo importantes para las consideraciones regulatorias y los 21 estándares de evidencia NICE para la evaluación y decisión de implementar herramientas de IA en el cuidado de la persona.

**Conclusiones:** Existen preocupaciones sobre la calidad de evidencia en la cual se realizan las ETES sobre herramientas de IA en salud. Por lo tanto, la aplicación de marcos de trabajo de NICE es recomendable para tomar la decisión estratégica de evaluación e implementación de IA. De igual manera, debemos promover estudios de alta calidad metodológica para la validación externa de herramientas de IA con el fin de mejorar la calidad de la evidencia y de las ETES posteriores.

**Palabras clave:** technology assessment, biomedical; implementation science; artificial intelligence

## Introducción

La Inteligencia Artificial (IA) es definida como una rama de la ciencia de computación, estadística e ingeniería que usa algoritmos o modelos para realizar tareas y exhibir comportamientos como el aprendizaje, toma de decisiones y realización de predicciones (1). Sus aplicaciones a la salud se han incrementado, desde el tamizaje de enfermedades hasta el tratamiento individualizado de patologías complejas.

Es una tecnología con un avance incremental que se visualiza en el número de publicaciones y cada día más aplicaciones en la salud con buenos rendimientos sobre sus aplicaciones y potenciales aplicaciones para mejorar y restaurar la salud de las personas (2). Sin embargo, sus grandes ventajas vienen con grandes preocupaciones sobre la validación de estas herramientas en el mundo real y en el mundo actual donde la seguridad de los datos es una preocupación diaria.

Por lo anterior, este informe técnico pretende emitir un concepto a la luz de la evidencia actual sobre la evaluación e implementación de la IA en el cuidado de las personas.

## Objetivo

Emitir un concepto basado en evidencia sobre la evaluación e implementación de la IA en el cuidado de las personas

## Alcance

Profesionales de la salud que en su práctica clínica vayan a evaluar e implementar la IA en el cuidado de las personas.

## Población

### 1.1. Población diana

Personas que reciben atención en salud, profesionales de la salud, desarrolladores de herramientas en IA, realizadores de política sobre inteligencia artificial en salud, evaluadores de tecnologías en salud.

### 1.2. Población no incluida

No se excluye ninguna persona del alcance de este documento

## Aspectos de salud y ámbito de aplicación

La IA es susceptible de aplicarse en el promoción de la salud y prevención de la enfermedad, diagnóstico, tratamiento y seguimiento, así como en el ámbito hospitalario, ambulatorio y comunitario.

## Usuarios

Tomadores de decisión en el cuidado de las personas.

## Metodología

El presente concepto basado en evidencia fue elaborado siguiendo el Manual para la elaboración de revisiones rápida de evidencia del Instituto Global de Excelencia Clínica (IGEC) Keralty.

Este documento responde a la pregunta: ¿Cuáles son las consideraciones en la evaluación e implementación de la IA en el cuidado de las personas?

Se realizó una búsqueda sistemática en MEDLINE a través de PubMed con la siguiente estrategia de búsqueda:

*((("technology assessment, biomedical"[MeSH Terms] OR ("technology"[All Fields] AND "assessment"[All Fields] AND "biomedical"[All Fields])) OR "biomedical technology assessment"[All Fields] OR ("health"[All Fields] AND "technology"[All Fields] AND "assessment"[All Fields])) OR "health technology assessment"[All Fields]*

OR ("evidence based medicine"[MeSH Terms] OR ("evidence based"[All Fields] AND "medicine"[All Fields]) OR "evidence based medicine"[All Fields] OR ("evidence"[All Fields] AND "based"[All Fields] AND "medicine"[All Fields]) OR "evidence based medicine"[All Fields]) OR ("evidence based practice"[MeSH Terms] OR ("evidence based"[All Fields] AND "practice"[All Fields]) OR "evidence based practice"[All Fields] OR ("evidence"[All Fields] AND "based"[All Fields] AND "practice"[All Fields]) OR "evidence based practice"[All Fields]) OR ("implementation science"[MeSH Terms] OR ("implementation"[All Fields] AND "science"[All Fields]) OR "implementation science"[All Fields]) OR (("ambulatory care facilities"[MeSH Terms] OR ("ambulatory"[All Fields] AND "care"[All Fields] AND "facilities"[All Fields]) OR "ambulatory care facilities"[All Fields] OR "clinic"[All Fields] OR "clinic s"[All Fields] OR "clinical"[All Fields] OR "clinically"[All Fields] OR "clinicals"[All Fields] OR "clinics"[All Fields]) AND ("guideline"[Publication Type] OR "guidelines as topic"[MeSH Terms] OR "guidelines"[All Fields]))) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields] OR ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) OR ("diagnosis, computer assisted"[MeSH Terms] OR ("diagnosis"[All Fields] AND "computer assisted"[All Fields]) OR "computer-assisted diagnosis"[All Fields] OR ("computer"[All Fields] AND "aided"[All Fields] AND "diagnosis"[All Fields]) OR "computer aided diagnosis"[All Fields]) OR (("artificial"[All Fields] OR "artificially"[All Fields]) AND "intelligence-based"[All Fields] AND ("equipment and supplies"[MeSH Terms] OR ("equipment"[All Fields] AND "supplies"[All Fields]) OR "equipment and supplies"[All Fields] OR ("medical"[All Fields] AND "device"[All Fields]) OR "medical device"[All Fields])) OR ("mayo clin proc digit health"[Journal] OR "lancet digit health"[Journal] OR "eur heart j digit health"[Journal] OR "digit health"[Journal] OR ("digital"[All Fields] AND "health"[All Fields]) OR "digital health"[All Fields])) AND ((guideline[Filter] OR practiceguideline[Filter] OR systematicreview[Filter]) AND (2018:2023[pdat]))

Se incluyeron aquellas publicaciones con diseño de revisiones sistemáticas o guías de práctica clínica con información sobre la evaluación e implementación de la inteligencia artificial en el cuidado de las personas. Se excluyeron estudios primarios, comentarios editoriales, cartas al editor y publicaciones mayores a 5 años.

Posteriormente, se hizo el tamizaje por títulos y resúmenes por medio de un epidemiólogo, en aquellos incluidos se accedió al texto completo para verificar condiciones de elegibilidad. En aquellos textos completos seleccionados se hizo la extracción de información que compone el cuerpo de evidencia del presente documento. Adicional a la búsqueda sistemática, se incluyeron documentos técnicos con relevancia para el alcance del presente documento.

## Resumen de evidencia y recomendaciones

Se encontraron 1460 registros de acuerdo a la estrategia de búsqueda planteada. Se revisaron títulos y resúmenes incluyendo 16 artículos para revisión de texto completo. Los anteriores fueron descargados para la revisión de texto completo, siendo elegidos para extracción 11 revisiones sistemáticas y 2 informes técnicos.

### Características de la evidencia

El cuerpo de evidencia correspondiente a revisiones sistemáticas fue publicado en el periodo de 2020 a 2023, con enfoques desde el cuidado general del paciente, pasando por la atención primaria, atención de enfermedades crónicas y la evaluación de calidad de estudios en el contexto de las evaluaciones de tecnología en salud. La evidencia se concentra en países desarrollados como Estados Unidos, Reino Unido, Unión Europea, Canadá, Australia donde hay más frecuencia de registro.

La evidencia incluye aplicaciones de IA en múltiples campos (Enfermedades metabólicas, circulatorias, respiratorias, cáncer, cuidado crítico, radiología, sepsis, telemedicina, neurología, cirugía general, entre otras) teniendo en cuenta la perspectiva de aplicación en el mundo real e implementación. Las áreas de aplicación de la IA pueden resumirse en 3 dominios: recolección de datos médicos, análisis de datos médicos y procesos de tratamiento activo (3).

### Limitaciones Metodológicas

Abdulazeem y colaboradores demuestran que la mayoría de estudios sobre herramientas de IA son sobre el desarrollo del modelo de predicción sin procesos de validación externa del modelo, y en estos estudios el 90,5% tenían un riesgo alto de sesgo (4). De igual manera, Ben Israel y colaboradores, resalta que la mayoría de estudios son de temporalidad retrospectiva y de calidad pobre que conlleva a limitaciones significativas en la implementación de la IA en la práctica clínica diaria (2).

Alguna evidencia logra mostrar estudios de implementación, por ejemplo, Yin y colaboradores muestra alto alcance de la IA dado sus aplicaciones de tamizaje y clasificación del riesgo, diagnósticos, análisis de riesgo y tratamiento, sin embargo, son de alta heterogeneidad y con tamaños de muestra pequeños de implementación (5). Se reportaron reducciones de costos con algunas herramientas en el contexto de sepsis (\$25.611 USD) e infecciones del trato respiratorio inferior (\$ 3.630 USD) en comparación al cuidado usual de estas enfermedades. Wolff y colaboradores buscaron evidencia sobre desenlaces de costo efectividad en la implementación de IA encontrando escasa producción bibliográfica con importantes deficiencias en la calidad de evaluación económica (6).

## Calidad de Evaluaciones de Tecnología Existentes

Algunos estudios encontrados se enfocaron en estudiar la calidad de los estudios sobre IA y su implicación en la realización de Evaluaciones de Tecnología en Salud (ETES). Farah y colaboradores encontraron que bajo marcos de trabajo internacionales, la mayoría de las ETES se enfocan al reporte de la eficacia, pero hay bajo reporte de seguridad, aspectos económicos y en ninguno de los casos se evaluó el componente ético (9). En otros estudios se encontraron limitaciones originadas en la heterogeneidad de desenlaces y aplicaciones, preocupaciones sobre la validación de los datos, y el impacto económico de la implementación de IA (10,11).

## Implementación de Herramientas IA

Una de las principales barreras para la implementación es el “black box dilema”, no entender la forma en la que funcionan las herramientas de IA para el soporte clínico, generando baja confiabilidad que empeora por la ausencia en la evidencia de guías metodológicas, detalle del abordaje de datos, interpretación y reproducibilidad de la herramienta (2,7,8).

Las barreras para la implementación pueden entenderse desde la óptica de desarrolladores de herramientas IA, profesionales de la salud, administradores en salud, realizadores de políticas públicas y pacientes (12). Desde el desarrollo de la herramienta por la complejidad de la misma, falta de interoperabilidad, disponibilidad y almacenamiento de los datos. Los profesionales de salud saben de sus ventajas en la eficiencia operacional y beneficios para el paciente pero dado la falta de conocimiento sobre su funcionamiento disminuyen la usabilidad (“black box dilema”).

Desde la óptica administrativa consideran su aplicación al desarrollo de tareas repetitivas pero hay desconfianza sobre su uso en la práctica clínica diaria. Los realizadores de políticas hacen énfasis en la necesidad de colaboración entre actores, resaltan que el contexto de implementación es variable aún dentro de un mismo hospital y hay preocupaciones sobre las responsabilidades legales en el uso de IA. Los pacientes ven como una preocupación la disminución de contacto con los profesionales de la salud.

## Marcos de trabajo para la adecuada evaluación e implementación

Con el fin de superar las barreras en la implementación de IA en la práctica clínica se encontró en la evidencia, dos importantes documentos que nos permiten entender aspectos regulatorios y consideraciones en la realización de ETES. Recientemente la Organización Mundial de la Salud (OMS) ha publicado un documento sobre las consideraciones regulatorias en el uso de IA en salud (13). Se centra en 6 puntos principales que se deben tener en cuenta en el momento de regulación de la implementación:

- 1. Documentación y transparencia:** Debe existir un protocolo que especifique el propósito de la herramienta y el desarrollo esperado del proceso. Lo anterior con el fin de poder identificar desviaciones dentro del proceso, y en caso de encontrarlas poder justificar dichas desviaciones.
- 2. Manejo del riesgo y desarrollo de sistemas basados en IA:** Debe considerarse el abordaje del todo el ciclo de vida de la herramienta (pre-mercadeo, manejo del desarrollo, vigilancia post-mercadeo y manejo del cambio). Debe existir una evaluación de los riesgos asociados desde la ciberseguridad, vulnerabilidades y posibles sesgos metodológicos.
- 3. Validación clínico – analítica e intención de uso:** La herramienta de IA debe ser validada en estudios prospectivos, idealmente bajo un ensayo clínico aleatorizado, o aquel que mejor se ajuste a la validación clínica requerida. De igual forma, similar a los medicamentos, debe existir una vigilancia estricta post-mercadeo en busca de situaciones que indiquen la necesidad del cambio de la herramienta.
- 4. Calidad del dato:** El rendimiento de la herramienta de IA depende directamente de datos de alta calidad, por lo tanto, se debe garantizar el desarrollo de la herramienta con altos estándares en la calidad del dato. De igual forma, durante la implementación, los datos generados deben tener la estructura para poder hacer nuevos entrenamientos de los modelos y asegurar adecuados procesos de validación posterior.
- 5. Protección y privacidad de datos:** Desde la planeación del desarrollo de la herramienta, se debe prever los sistemas y aspectos regulatorios necesarios para asegurar la protección y privacidad de los datos. En caso de vulnerabilidad, debe existir un proceso claro sobre cómo proceder en caso de riesgos latentes a la seguridad de los datos.
- 6. Colaboración y compromiso:** Las herramientas de IA requiere la participación de varios actores, para lograrlo, se debe procurar procesos y plataformas que lleven a la armonía entre los actores que permitan el mejor desarrollo e implementación de la herramienta.

## **Estandares de Evidencia para la Evaluación e Implementación de Herramientas IA en el cuidado de las personas**

Con el fin de dar claridad y practicidad al proceso de desarrollo hasta la implementación, NICE ha publicado el marco de trabajo para establecer los estándares de evidencia específica para las tecnologías digitales donde encontramos las herramientas de IA (14). De forma inicial, deja claro hacia que tecnologías va dirigida dicho marco de trabajo:

- Aplicaciones de smartphones o softwares

- Herramientas en línea para tratar o diagnosticar condiciones, prevenir la enfermedad o para mejorar la eficiencia operacional
- Programas que pueden ser usados para analizar datos de dispositivos.

Las tecnologías de salud digital, incluidas aquellas con IA, son clasificadas por NICE en 3 niveles:

**Nivel A:** Aquellas que van dirigidas a ahorrar costos o mejorar procesos, no tienen contacto directo con el paciente ni desenlaces de salud.

**Nivel B:** Aquellas que ayudan a las personas a manejar su propia salud y bienestar

**Nivel C:** Aquellas que tratan y diagnostican condiciones o guían decisiones del cuidado de las personas.

Dicha clasificación es importante, dado que cada estándar de evidencia no aplica para todos los niveles. NICE contempla 21 estándares estratificados en 5 grupos de acuerdo al ciclo de vida del dispositivo

*Tabla 1. Estándares de evidencia NICE para la evaluación de tecnologías en salud en el ámbito de salud digital.*

Factor	Estándar	Nivel que aplica
<b>Diseño</b>	La herramienta debe cumplir con los estándares de seguridad y calidad	A, B y C
	Incorpora la intención de uso y su aceptabilidad en el diseño de la herramienta	A, B y C
	Considera la sostenibilidad ambiental	A, B y C
	Considera las inequidades del cuidado	A, B y C
	Incluye buenas prácticas en el diseño de la herramienta	A, B y C
	Define el nivel de involucramiento del profesional de salud	A, B y C
	Delimita procesos para crear información en salud confiable	B y C
	Muestra credibilidad entre los profesionales	B y C
	Da garantías en caso de que la herramienta se considere usar en grupos	B y C

	vulnerables (Niños o adultos en riesgo, según NICE)	
<b>Descripción de valor</b>	Describe la intención de uso y la población objetivo	A, B y C
	Describe el proceso del sistema	A, B y C
	Describe la vía propuesta usando la herramienta de IA	A, B y C
	Describe los impactos esperados en salud, costos y recursos comparados con el estándar o cuidado usual	A, B y C
<b>Demostración del rendimiento</b>	Da evidencia de la efectividad	C
	Evidencia del mundo real que muestra los beneficios en la práctica diaria	A, B y C
	La compañía y evaluar deben acordar un plan para medir usabilidad y potenciales cambios en el tiempo	A, B y C
<b>Entrega de valor</b>	Provee un análisis de impacto económico	A, B y C
	Para herramientas de alto riesgo financiero, se provee un análisis de costo-efectividad	A, B y C
<b>Consideraciones de implementación</b>	Asegura transparencia sobre requerimientos en la implementación	A, B y C
	Describe estrategias para comunicación, consentimiento y procesos de entrenamiento que permiten a la herramienta ser entendida por los usuarios finales	A, B y C
	Asegura escalabilidad apropiada	A, B y C

Estos dos marcos de trabajo, proveen los estándares de calidad y regulatorios que deben ser evaluados antes de la inclusión de herramientas de IA en el cuidado de las personas.

## **Conclusión y Recomendaciones**

Las herramientas de IA están en auge y se espera un crecimiento mayor en los siguientes años, a la fecha, la evidencia es heterogénea y con altas preocupaciones sobre su calidad dado la presencia de sesgos, ausencia de validación externa, sin evaluaciones del impacto económico y ético. Por lo anterior, la decisión de implementar la IA se realiza sobre incertidumbres en la calidad y certeza de la evidencia.

Sin embargo, marcos de trabajo como el de la OMS y NICE ofrecen los estándares a cumplir la evidencia para apoyar la decisión estratégica de implementar una herramienta de IA en el cuidado de las personas, proveniente de una evaluación correcta de la tecnología. Este informe técnico propone adoptar los estándares de evidencia NICE para apoyar la toma de decisiones sobre la evaluación e implementación de dispositivos de IA en el cuidado de las personas. De igual manera, es importante resaltar la necesidad de realizar estudios sobre la efectividad en mundo real y la implementación de este tipo de herramientas con el fin de describir el valor generado en el cuidado de las personas.

## Bibliografía

1. IMDRF. Machine Learning-enabled Medical Devices-A subset of Artificial Intelligence-enabled Medical Devices: Key Terms and Definitions Authoring Group: IMDRF AIMD Working Group. 2021.
2. Ben-Israel D, Jacobs WB, Casha S, Lang S, Ryu WHA, de Lotbiniere-Bassett M, et al. The impact of machine learning on patient care: A systematic review. Vol. 103, Artificial Intelligence in Medicine. Elsevier B.V.; 2020.
3. Bitkina OV, Park J, Kim HK. Application of artificial intelligence in medical technologies: A systematic review of main trends. Vol. 9, Digital Health. SAGE Publications Inc.; 2023.
4. Abdulazeem H, Whitelaw S, Schauburger G, Klug SJ. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data. PLoS One. 2023 Sep 1;18(9):e0274276.
5. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: Systematic review. Vol. 23, Journal of Medical Internet Research. JMIR Publications Inc.; 2021.
6. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: Systematic review. Vol. 22, Journal of Medical Internet Research. JMIR Publications Inc.; 2020.
7. Kolasa K, Kozinski G. How to value digital health interventions? A systematic literature review. Vol. 17, International Journal of Environmental Research and Public Health. MDPI; 2020.
8. Yusuf M, Atal I, Li J, Smith P, Ravaud P, Fergie M, et al. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. BMJ Open. 2020 Mar 23;10(3):e034568.
9. Farah L, Davaze-Schneider J, Martin T, Nguyen P, Borget I, Martelli N. Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A systematic review. Vol. 140, Artificial Intelligence in Medicine. Elsevier B.V.; 2023.
10. Von Huben A, Howell M, Howard K, Carrello J, Norris S. Health technology assessment for digital technologies that manage chronic disease: A systematic review. International Journal of Technology Assessment in Health Care. Cambridge University Press; 2021.
11. Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah. Value in Health. 2022 Mar 1;25(3):340–9.

12. Al-Zubaidy M, Hogg HJ, Maniatopoulos G, Talks J, Teare MD, Keane PA, et al. Stakeholder Perspectives on Clinical Decision Support Tools to Inform Clinical Artificial Intelligence Implementation: Protocol for a Framework Synthesis for Qualitative Evidence. JMIR Res Protoc. 2022 Apr 1;11(4).
13. World Health Organization. Regulatory considerations on artificial intelligence for health. 2023.
14. NICE. Evidence standards framework for digital health technologies [Internet]. 2022. Available from: [www.nice.org.uk/corporate/ecd7](http://www.nice.org.uk/corporate/ecd7)

## Anexo 1. Tabla de Evidencia

Autor	Año Publicación	Enfoque	Años Búsqueda	Países	Enfermedades	Resultados
Abdulazeem	2023	Atención Primaria	1990 - 2022	USA, UK (57,1%)	Metabólicas, mentales, circulatorias, respiratorias, cáncer. (n = 42)	La mayoría de estudios incluían desarrollo de modelos de predicción (n = 207) sin validación externa (76.5%). El 25.5% era en enfermedades endocrinas, nutricionales y enfermedades metabólicas. El 90.5% de los estudios incluidos tenían riesgo de seso alto o no claro (PROBAST)
Ben-Israel	2020	Cuidado general del paciente	2000 - 2018	No específica	Oncología (17%), neurología, cardiología, cuidado crítico, neurocirugía, cirugía general.	Los estudios incluidos eran clasificados de calidad pobre, ninguno fue considerado de buena calidad. La gran mayoría de los artículos fueron retrospectivos. Existen limitaciones significativas en la implementación de la IA en la práctica clínica diaria. Desconocimiento del funcionamiento de las metodologías IA (Black box dilema).
Bitkina	2023	General	2011 - 2021	USA, Canadá, UK	Oncología, neumología, medicina cardiovascular, ortopedia, hepatología y neurología.	<p>Recolección de datos médicos, análisis de datos médicos y procesos de tratamiento activo.</p> <p>IA en medicina en 3 áreas: Recolección de datos (detección, extracción), análisis de enfermedad (Clasificación, Identificación), tratamiento activo (Predicción, pronostico)</p> <p>Se reconocen limitaciones y sesgos en la aplicación en la vida real. Verificación de los datos, calidad de los datos validadores, altos costos de dispositivos con elementos de IA. Replicación limitada de los estudios ante la complejidad metodológica.</p>

Farah	2023	Calidad de estudios clínicos para soportar evaluaciones de tecnología	2016 - 2021	Europa (43%), Asia (28,5%), Norte América (25%), Australia (3,5%)	Radiología	<p>Estudios retrospectivos (89%), prospectivos (11%), ECA (n = 2).</p> <p>MI-CLAIM, media 67%; &gt;70 % (n=18, 32%); &gt; 50 % &lt; 70% (n=28, 50%); &lt; 50% (n = 10, 18%)</p> <p>Diseños de estudios, optimización de datos, rendimiento del modelo (&gt;60%). Reproducibilidad y evaluación en clínica (&lt;30%).</p> <p>HTA - Domains, media 48%; &gt;70 % (18%); &gt; 50 % &lt; 70% (29%); &lt; 50% (53%). Solo en 9% se evaluó seguridad, solo en 20% los aspectos económicos, ningún artículo menciona componentes éticos.</p> <p>Recomendaciones: Mejorar vigilancia post-mercadeo, Evaluar transparencia, interpretabilidad, explicabilidad, componente ético.</p>
-------	------	---	-------------	---	------------	---

Hogg	2023	Perspectiva de los actores clave en la implementación clínica de la IA	2014 - 2021	23 países (Estados Unidos, Reino Unido, Canadá y Australia)	25 especialidades (dominancia de la atención primaria)	Se identifican 5 grupos de importancia: Desarrolladores (Se detiene implementación por la complejidad de la herramienta, no existe experiencia, falta de interoperabilidad, disponibilidad y almacenamiento de los datos), profesionales de la salud (potencial herramienta para facilitar el entrenamiento clínica, tareas simples o repetitivas, escepticismo sobre la habilidad de comunicación, desconocimiento de los métodos de la IA), managers y lider (Se prioriza herramientas de IA que mejoren procesos y ayuden a proteger la inversión, sin embargo, hay relaciones negativas ante el alcance de estas herramientas), pacientes y público general (Preocupaciones sobre el reemplazo del profesional de salud en áreas como salud mental o trabajo social, se perciben mejor en la enfermedad crónica, incluir la perspectiva del paciente en el desarrollo de las herramientas), realizadores de políticas y reguladores (Necesidad de colaboración entre actores, especificidades de cada lugar e implementación, responsabilidades legales de la aplicación clínica)
Voets	2023	Calidad de las evaluaciones de tecnología en salud sobre aplicaciones de IA en el ámbito de la salud	2016 - 2021	No específica	Oftalmología una de las principales	Se identifican limitaciones cualitativas y cuantitativas, métodos heterogéneos de evaluación en cada tecnología. Media de calidad CHEERS 66% (Puntajes bajos en validación, caracterización de heterogeneidad)

von Huben	2021	Enfermedad Crónicas	2015 - 2020	Países Unión Europea, Reino Unido, Asia Pacifico	Salud Digital, eHealth, mHealth	Los dominios de efectividad clínica y seguridad fueron los más cubiertos. Problemas en la validación económica de los modelos, no se cubrió casi el aspecto de pacientes y vista social. Aplicación de los estándares de NICE. Propone requerimientos adicionales en las HTA's de intervenciones de salud digital (Tabla3)
Yin	2021	IA aplicaciones en vida real	2010 - 2020	Estados Unidos (n=20), UK (n=5), Australia (n=2), Canadá, Japón, Alemania (n=1), Israel, España.	Sepsis, cáncer de mama, retinopatía diabética, adenomas y pólipos, cataratas, evento cerebrovascular	Tamizaje y clasificación del riesgo, diagnostico, análisis de riesgo, tratamiento. Alta heterogeneidad de los resultados. pocos estudios de implementación con pequeños tamaños de muestra. Costos en un estudio para la vigilancia de antibióticos reduce costos por 25,611 USD para sepsis y 3,630 USD para infecciones del tracto respiratorio inferior en comparación al cuidado usual.
Kolasa	2020	Evaluaciones de tecnología de intervenciones con Salud Digital	1998 - 2019	Estados Unidos, Canadá, España, Reino Unido	Telemedicina, Salud Móvil	Escasas descripciones del problema de salud o necesidades de salud, involucramiento de los actores, no hay estándar en el cálculo del impacto económico.
Yusuf	2020	Calidad de estudios que desarrollan y validan modelos de IA	2015 - 2018	No específica	Oncología, neurología	Los estudios encontrados no reportan guías metodológicas para su ejecución, así como la falta detalle del abordaje de datos, interpretación y reproducibilidad
Wolff	2020	Evaluación de calidad de estudios de costo-efectividad sobre implementación de AI en el cuidado de la salud	Últimos 5 años desde el momento de búsqueda	No específica	No específica	Escasa producción bibliográfica sobre estudios de costo efectividad en la implementación de IA en el cuidado de la salud. Se reportan importantes deficiencias en la calidad de evaluación económica en los estudios reportados.