

VALIDACIÓN E IMPLEMENTACIÓN DE IA EN SALUD

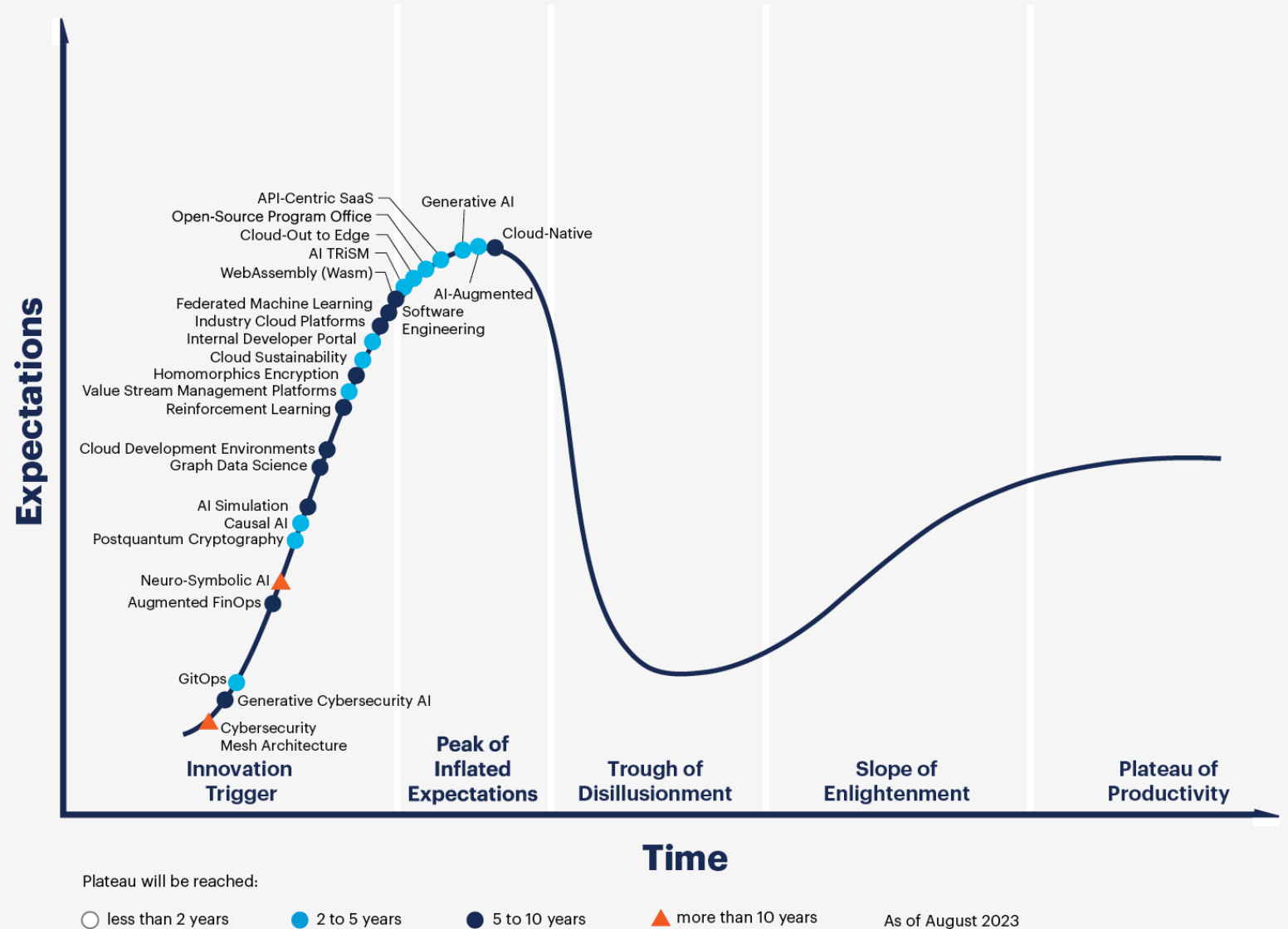
Dr. Stefan Escobar
Agreda
Dirección de Telemedicina
Unidad de Telesalud





Hype Cycle for Emerging Technologies, 2023

¿Será tan buena la IA como parece?



Consideraciones Regulatorias en Inteligencia Artificial en Salud

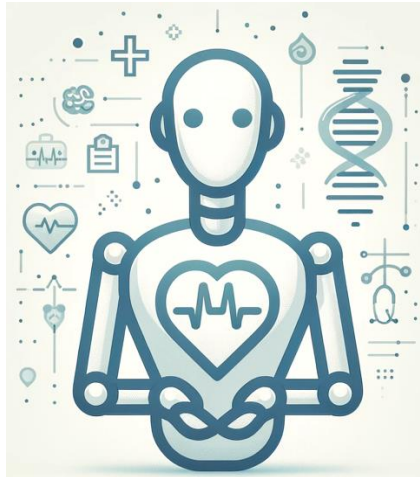
(Organización Mundial
de la Salud)

Organization WH. Regulatory considerations on artificial intelligence for health [Internet]. World Health Organization; 2023



¿Cómo se regula la Inteligencia Artificial en Salud?

IA (Software)



Dispositivo biomédico

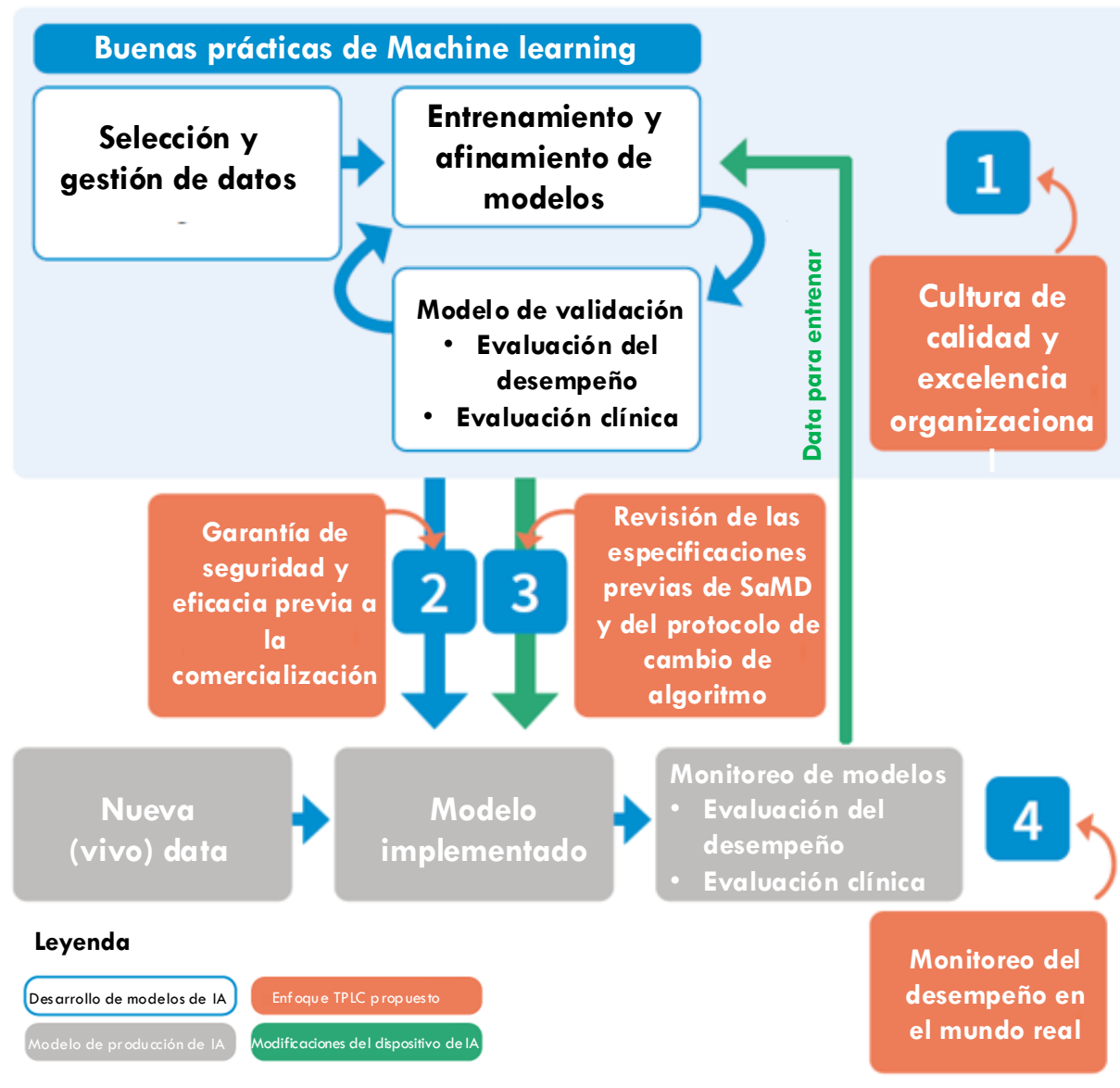


***SaMD** Software as a Medical Device

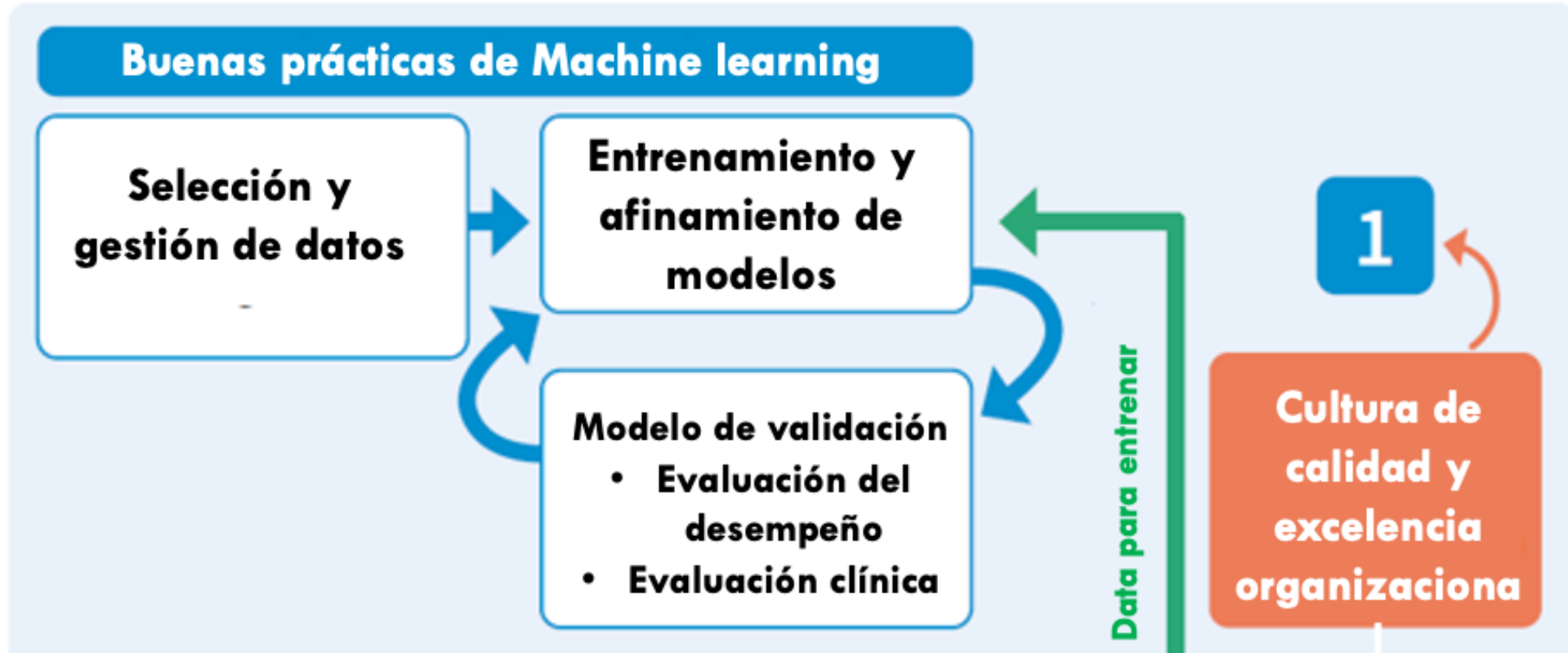


*SaMD Software as a Medical Device

Enfoque del Ciclo de Vida Total del Producto (TPLC) de un sistema de IA



Organization WH. Regulatory considerations on artificial intelligence for health [Internet]. World Health Organization; 2023

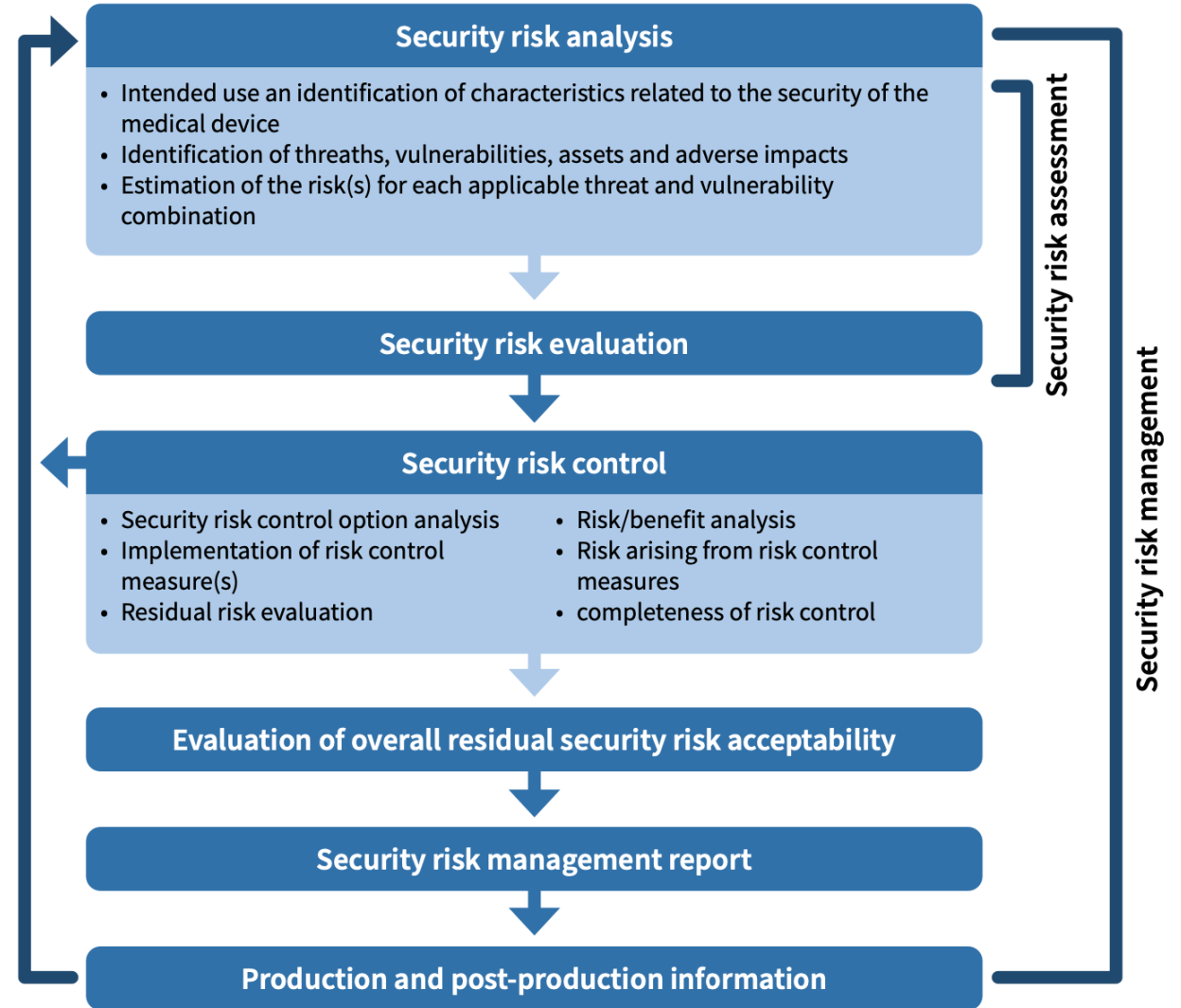
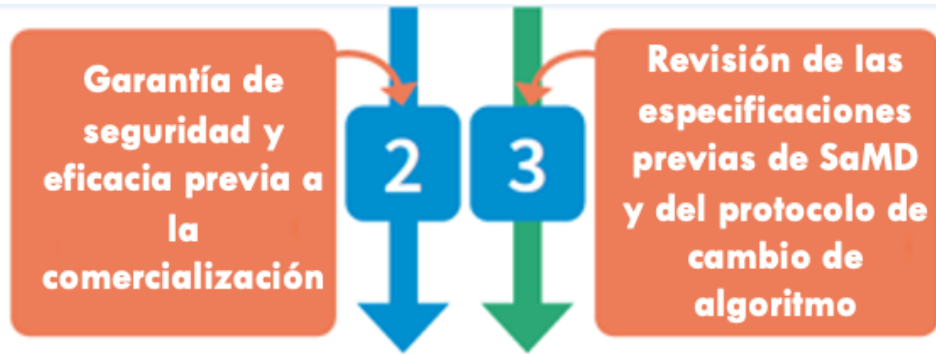


Principios de Buenas Prácticas en Machine Learning (FDA)

Health C for D and R. Good Machine Learning Practice for Medical Device Development: Guiding Principles. FDA [Internet].



Verificación de la Seguridad del Modelo de IA desarrollado



Clasificación de RIESGO de los sistemas de IA

Estado de situación o condición clínica	Implicancia de la información proporcionada por el sistema de IA en la decisión clínica		
	Tratamiento o diagnóstico	Impulsar la gestión clínica	Informar gestión clínica
Crítico	IV	III	II
Serio	III	II	I
No serio	II	I	I

Clasificación de RIESGO de los sistemas de IA

Clase I: Riesgo bajo

- No está destinado a dar soporte o mantener la vida.
- No presentan un riesgo potencial de enfermedad o lesión.
- Uso generalmente simple.

Para FDA:

- No requieren estudios previos.
- Generalmente exentos de notificación previa (510(k)).
- Deben cumplir con buenas prácticas de manufactura.

EJEMPLOS

- ✓ Monitores de actividad física.
- ✓ Monitores de sueño.



Clasificación de RIESGO de los sistemas de IA

Clase II: Riesgo moderado

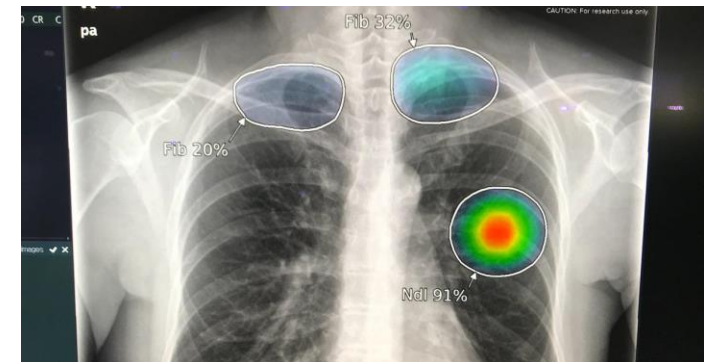
- Los controles generales por sí solos no pueden asegurar su seguridad y eficacia.
- Contacto más prolongado con pacientes.

Para FDA:

- Requiere notificación previa (510(k)): Debe demostrar equivalencia con la rutina o dispositivos aprobados.
- Controles generales y especiales: etiquetado, rendimiento, vigilancia post-comercialización, etc.

EJEMPLOS

- ✓ Dispositivos para asistir la interpretación de imág. médicas.
- ✓ Software para el triaje de pacientes según nivel de urgencia.



Clasificación de RIESGO de los sistemas de IA

Clase III: Riesgo alto

- Sostienen o soportan la vida. ó
- Son implantables. ó
- Presentan un riesgo potencial irrazonable de enfermedad o lesión.

Para FDA:

- Requiere aprobación previa (PMA), con estudios clínicos que demuestren su seguridad y eficacia.
- Exige evidencia científica robusta.

EJEMPLOS

- ✓ Monitor de funciones vitales.
- ✓ Implante para monitoreo de glucosa.






Clasificación de RIESGO de los sistemas de IA

New Search Help | More About 21CFR

[Code of Federal Regulations]
[Title 21, Volume 8]
[CITE: 21CFR892.2020]



TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER H - MEDICAL DEVICES

PART 892 -- RADIOLOGY DEVICES
Subpart B - Diagnostic Devices

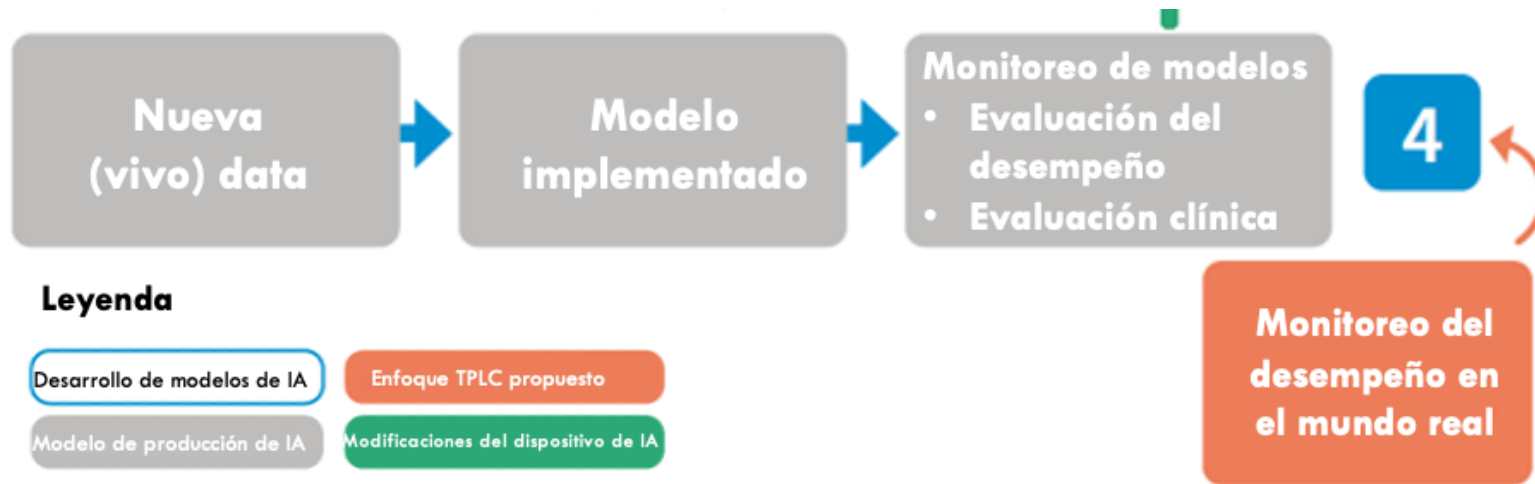
Sec. 892.2020 Medical image communications device.

(a) *Identification.* A medical image communications device provides electronic transfer of medical image data between medical devices. It may include a physical communications medium, modems, and interfaces. It may provide simple image review software functionality for medical image processing and manipulation, such as grayscale window and level, zoom and pan, user delineated geometric measurements, compression, or user added image annotations. The device does not perform advanced image processing or complex quantitative functions. This does not include electronic transfer of medical image software functions.

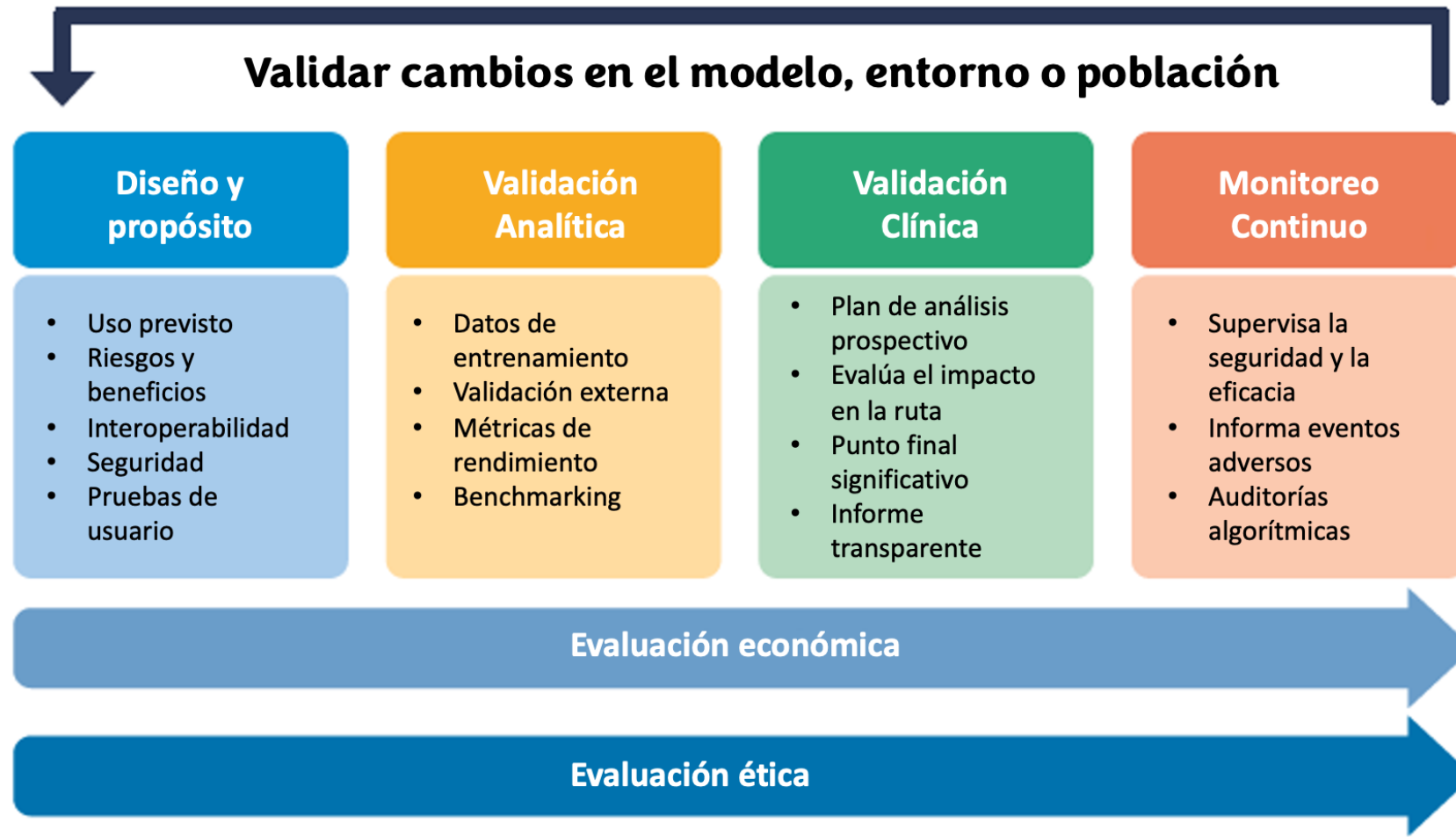
(b) *Classification.* Class I (general controls). The device is exempt from the premarket notification procedures in subpart E of part 807 of this chapter subject to § 892.9.

[63 FR 23387, Apr. 29, 1998; 63 FR 44998, Aug. 24, 1998, as amended at 65 FR 2323, Jan. 14, 2000, 86 FR 20283, Apr. 19, 2021]

Implementación y monitoreo de modelos de IA



Marco para la evaluación clínica de modelos de IA

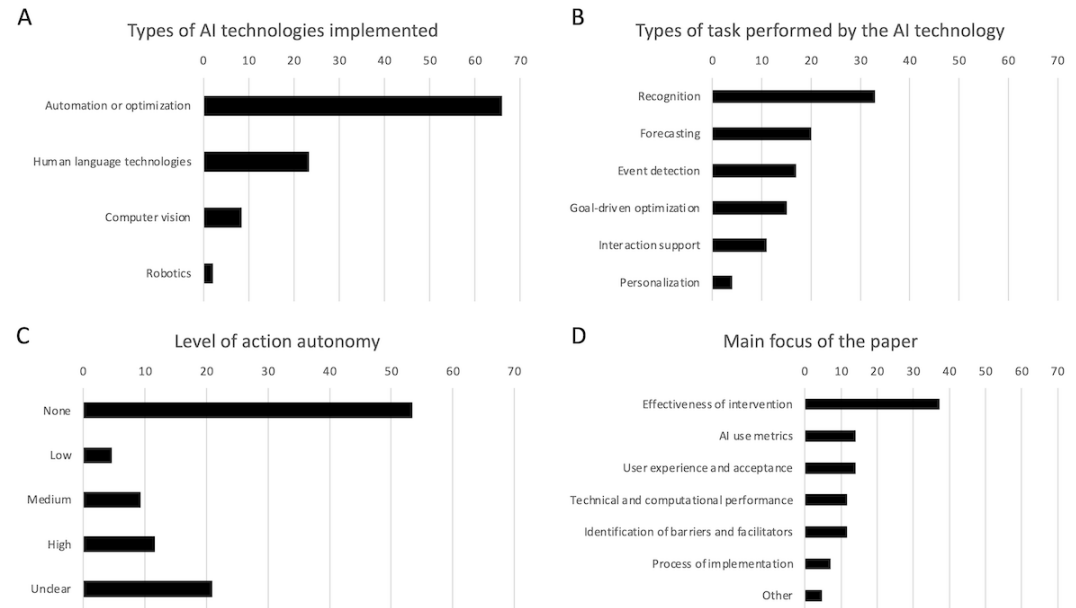
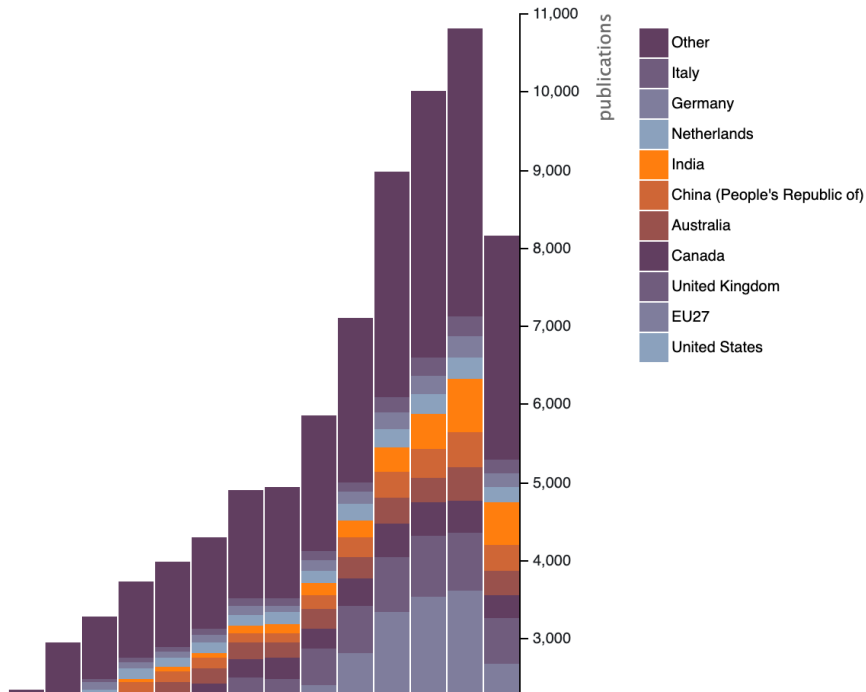


Organization WH. Regulatory considerations on artificial intelligence for health [Internet]. World Health Organization; 2023

Investigación vs implementación en IA

Miles de estudios en IA

Solo 40 han reportado su implementación hasta el 2023





Validación externa de Modelos de IA

Incluso los modelos de IA aprobados por FDA carecen de una adecuada validación externa

TABLE 3. Size of Validation Datasets Across Different Types of FDA-regulated AI/ML Algorithms. A few Algorithms had Either no Patient Evaluation or no Disclosure or Mention of Details Pertaining to Evaluation Patient Data

	Unclassified	CADt	CADe	CADx	CADe/x	IPQ	Total
No patient evaluation	4	0	1	0	0	12	17
<500 patients	6	15	3	3	5	13	45
500-1000 patients	1	6	0	0	1	2	10
>1000 patients	0	4	0	1	1	5	11
No available data	5	2	1	0	0	27	35
Total	16	27	5	4	7	59	118

AI/ML, artificial intelligence/machine learning; CAD, computer-aided; CADe, detection; CADe/x, detection/diagnosis; CADt, triage; CADx, diagnosis; FDA, Food and Drug Administration; IPQ, image processing/quantification.

¿Puedo estar seguro que una IA es buena porque tiene buenos resultados de predicción?

Depende también del tipo del tipo de análisis que se desarrolló para obtener estos resultados



Si basura entra, basura sale

¿Cómo puedo saber si el análisis de un modelo de IA fue desarrollado adecuadamente?



Niveles de evaluación/ Niveles de investigación en IA

Table 1. Translations of Stead et al’s levels of evaluation taxonomy to the research evaluation stage-based terminology.

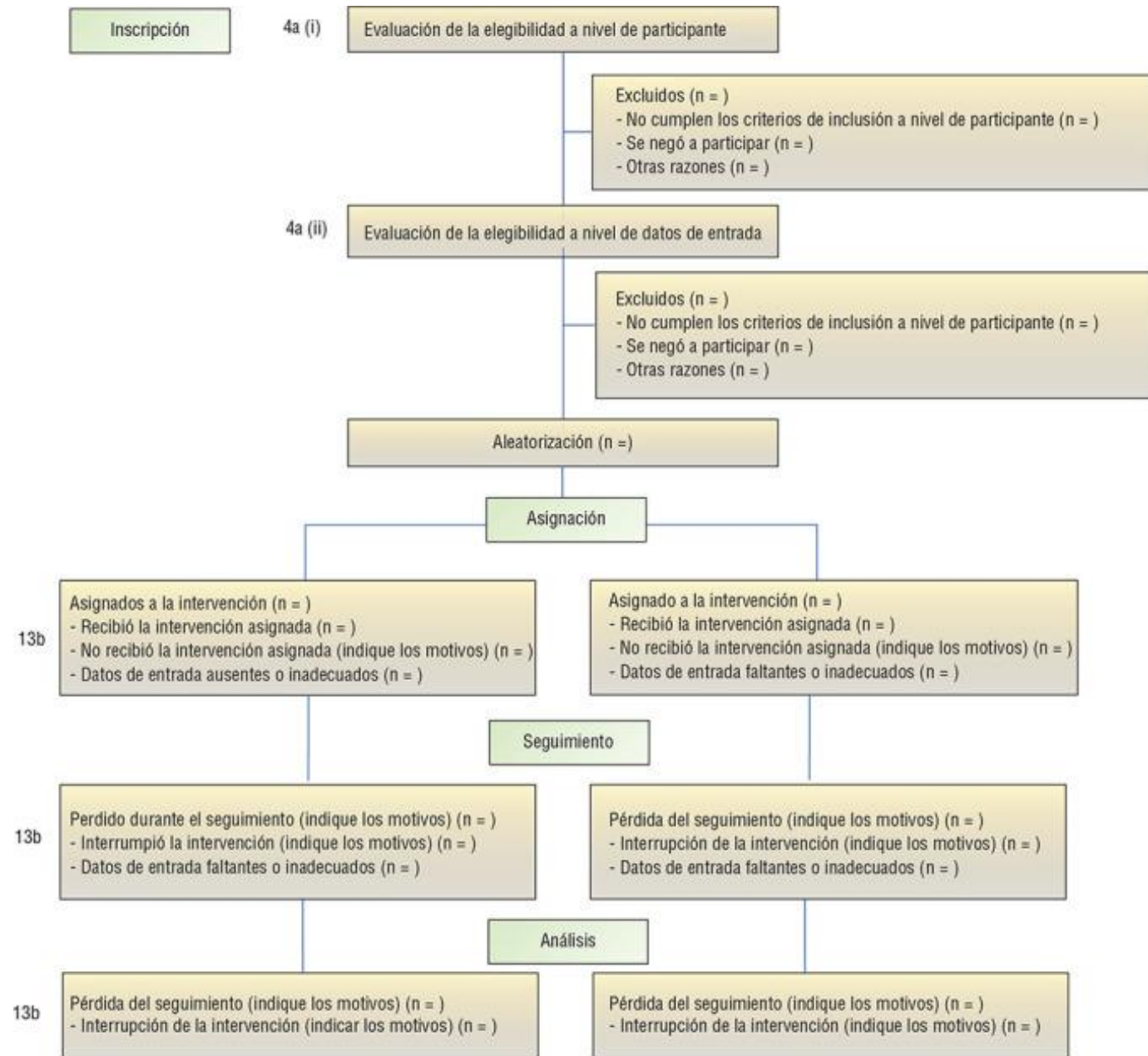
Implementation stage	Stead et al’s stage (level of evaluation)	Research evaluation stage and associated reporting standard
I	I. Definition	Definition (from Stead et al)
II	II. Lab-bench	Retrospective study (TRIPOD) ^{2,19}
III	III. Lab-field	Shadow/silent study (TRIPOD) ^{2,19}
IV	IV. Field	Small trial/pilot (DECIDE-AI) ²¹
V	V. Broader field	Large trial/RCT (CONSORT-AI) ^{22,43}

RCT: randomized controlled trial.

¿Cómo realizar un Ensayo Clínico Aleatorizado en IA?

CONSORT-AI (Resultados de estudio)

Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., Chan, A. W., ... & Rowley, S. (2024). Directrices para presentación de informes de ensayos clínicos sobre intervenciones con inteligencia artificial: extensión CONSORT-AI. Revista Panamericana de Salud Pública, 48, e13.





Modelo SALIENT para implementación de la IA

Anton H van der Vegt, Ian A Scott, Krishna Dermawan, Rudolf J Schnetler, Vikrant R Kalke, Paul J Lane, Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework, Journal of the American Medical Informatics Association, Volume 30, Issue 9, September 2023, Pages 1503–1515, <https://doi.org/10.1093/jamia/ocad088>

Revised SALIENT framework: components and tasks

Stages II and III Stage IV Stage V

Framework element A: specifications

Preparation—data pipeline

- PDP01 Identify, collect, and prepare data for development and validation, including training data^{54,58–60}
- PDP02 Establish interoperability and align to clinical coding standards (eg, ICD codes)^{33,49,59}
- PDP03 Investigate and establish IT hardware and storage capability⁶⁰

Preparation—AI model

- PAM01 Search for and evaluate existing AI models^{33,42}
- PAM02 Perform a cost benefit analysis and feasibility assessment of using AI^{27,35,53,59}

Preparation—clinical workflow

- PCW01 Create a plan to evaluate the success of the implementation^{38,53}

Framework element B: component development

Data pipeline (DP)

- DP11 Stress test infrastructure⁵⁹
- DP12 Scalability assessment: assess changes to data sources and input protocols^{27,35}

Artificial intelligence model (AM)

- AM05 Devise means and document interpretability of AI model outputs^{38,48,49,53,59}
- AM06 Define, publish, and update AI model fact label defining standardized communication of AI model information to end users^{33,38,57–59}
- AM07 Externally validate the model and assess generalizability^{33,49,53,58,59}
- AM08 Check for, report on, and apply methods to reduce overfitting⁵⁹
- AM09 Stress test AI model software⁵⁹

Human-computer interface (HC)

- HC5 Stress test HC interface software⁵⁹

Framework element E: routine use

Evaluation and monitoring

- EM12 Monitor and track data shift and AI quality^{27,33,35,50,53,54,58,59}
- EM13 Log AI decisions for traceability⁵⁹

Task	Stages II and III	Stage IV	Stage V
DP11	X	X	X
DP12		X	X
AM05	X	X	
AM06	X	X	X
AM07	X	X	
AM08	X	X	
AM09		X	
HC5		X	
EM12			X
EM13		X	X

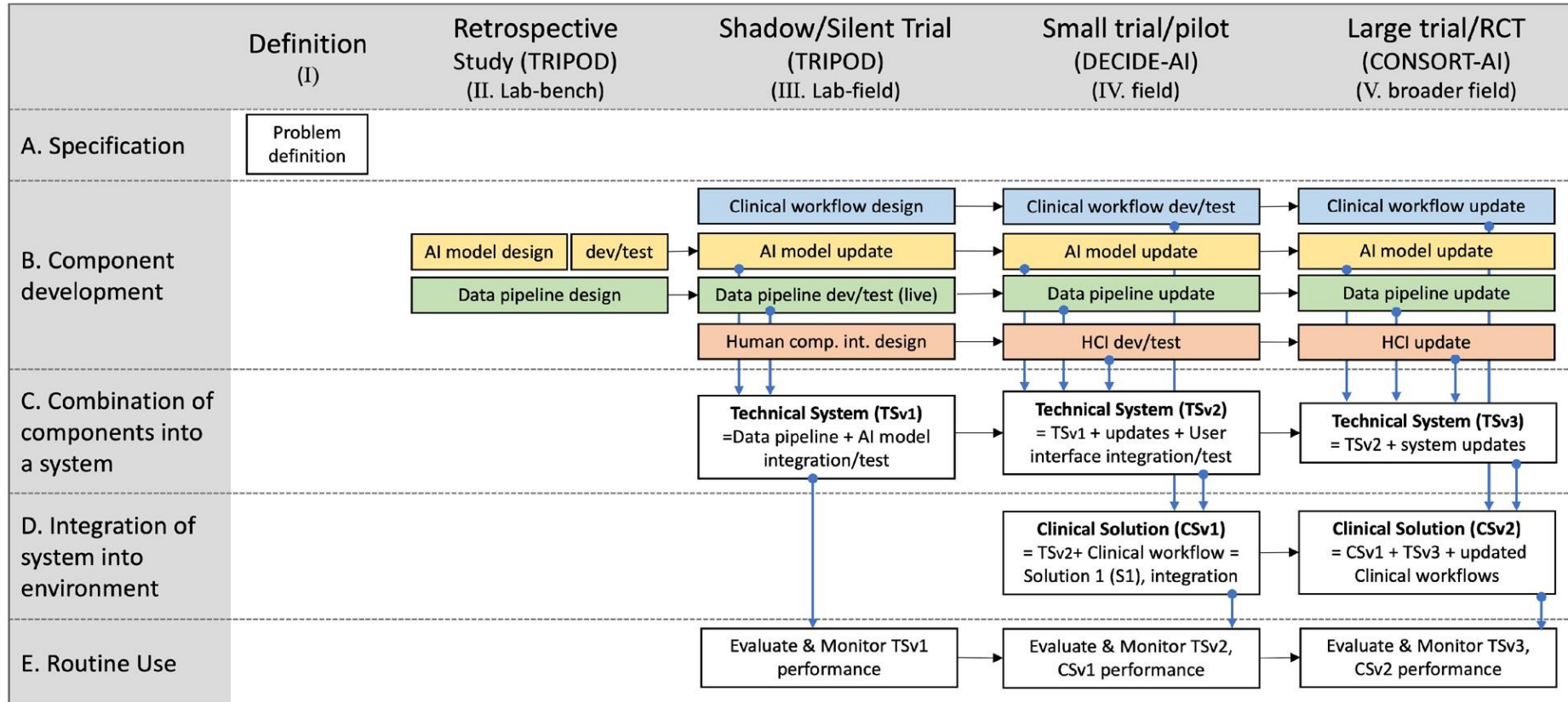
Applicable stages for each task, according to the TRIPOD, DECIDE-AI, and CONSORT-AI defined stages, are marked with an “x” in the appropriate columns, ie, for stage II/III (blue column), stage IV (amber column), and stage V (green column). Note that preparation tasks (PDP, PAM, PCW) are only applicable to stage I (Definition), which is not marked here.

ICD: International Classification of Diseases; IT: information technology; AI: artificial intelligence.

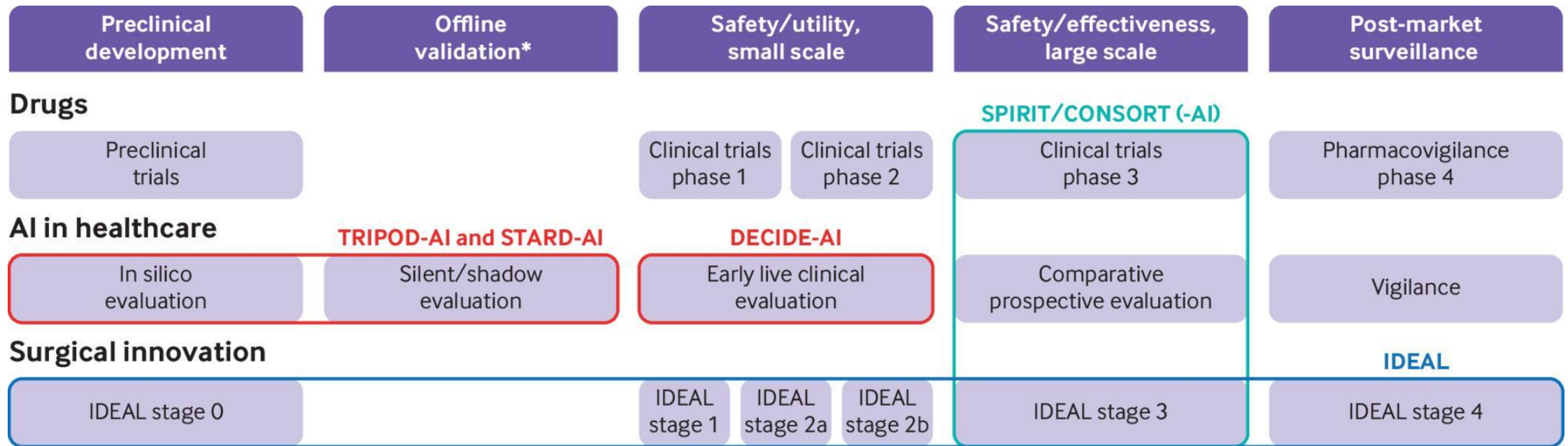
Modelo SALIENT para implementación de la IA



Relación de Modelo SALIENT con Niveles de Evaluación



Evaluación de medicamentos vs modelos de IA



Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., ... & McCulloch, P. (2022). Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *bmj*, 377.

Conclusiones

- Pese al enorme potencial de la IA en el sector salud, debe definirse **utilidad y alcance** de cada algoritmo en un **contexto real**.
- Como dispositivo biomédico la IA debe cumplir con **estándares de calidad mínimos** previo a considerar su implementación en campo.
- Las IA asociadas al sector clínico deben incorporar desde su desarrollo **estudios** que respalden su utilidad los cuales deben estar diseñados sobre la base de las **guías internacionales**.
- **Los requerimientos** de evaluación y evidencia requerida para cada modelo de IA dependen de su clasificación basado en su **nivel de riesgo**.

GRACIAS

