AEQUITAS

Regulación de la inteligencia artificial en la sanidad pública para prevenir los sesgos de género y raciales

Un modelo regulador de la IA biomédica acorde con la Carta Europea de Derechos Humanos

Documento de políticas

Co-funded by
the European Union

AEQUITAS

Introducción

La inteligencia artificial (IA) se integra cada vez más en los sistemas sanitarios de toda Europa, donde presta apoyo al diagnóstico, la toma de decisiones clínicas, la planificación de tratamientos y la predicción del riesgo de los y las pacientes. Estas tecnologías ofrecen oportunidades considerables para mejorar la eficiencia y los resultados de la asistencia sanitaria. Al mismo tiempo, cada vez hay más pruebas de que **los sistemas de IA pueden reproducir o amplificar las desigualdades existentes en la asistencia sanitaria**.[1]

Existen varios mecanismos que pueden dar lugar a resultados sesgados en la toma de decisiones asistida por IA en el ámbito sanitario. Entre ellos se incluyen los desequilibrios en los datos de entrenamiento, los errores de medición en las variables clínicas, el uso de indicadores sustitutivos inadecuados y las diferencias entre los entornos de desarrollo y los entornos de implementación en el mundo real.[2] El sesgo también puede surgir cuando los sistemas de IA se introducen en los flujos de trabajo clínico sin una formación adecuada para profesionales de la salud o sin mecanismos sustanciales de supervisión humana y rendición de cuentas que controlen su rendimiento en poblaciones de pacientes diversas.[3]

Además, la investigación empírica ha demostrado que dichos **sesgos pueden tener graves consecuencias en contextos clínicos**. Diversos estudios han señalado disparidades persistentes en el diagnóstico y el tratamiento de afecciones graves. Mujeres y pacientes pertenecientes a minorías raciales o étnicas con enfermedades cardiovasculares suelen recibir un diagnóstico y un tratamiento tardíos o menos intensivos en comparación con los hombres y los pacientes de raza blanca,[4] y las diferencias raciales y étnicas afectan a la atención y los resultados de la diabetes,[5] mientras que los trastornos de salud mental, como la depresión, están

---

[1] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y., *Artificial intelligence in healthcare: past, present and future*, *Stroke and Vascular Neurology* 2(4): 230–243, 2017.

[2] Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S., *Mitigating bias in machine learning for medicine*, *Communications Medicine* 1, 25, 2021.

[3] Panesar, S., Kliot, M., Parrish, R., Fernandez-Miranda, J., Cagle, Y., & Britz, G.W., *Promises and Perils of Artificial Intelligence in Neurosurgery*, *Neurosurgery* 87(1): 33–44, 2020.

[4] Balla, S., Gomez, S.E., & Rodriguez, F., *Disparities in Cardiovascular Care and Outcomes for Women From Racial/Ethnic Minority Backgrounds*, *Current Treatment Options in Cardiovascular Medicine* 22(12):75, 2020.

[5] Young, C. & Myers, A. K., *Racial and Ethnic Disparities in Diabetes Clinical Care and Management: A Narrative Review*, *Endocrine Practice* 29(4): 295–300, 2023.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

infradiagnosticados e infratratados tanto en las mujeres como en las poblaciones minoritarias.[6] Un ejemplo bien documentado es un estudio exploratorio publicado en el **Journal of Medical Internet Research**, que reveló disparidades de rendimiento basadas en el sexo en los algoritmos de aprendizaje automático utilizados para la predicción de enfermedades cardiacas; en concreto, que las pacientes mujeres estaban sistemáticamente infrarrepresentadas en los conjuntos de datos de entrenamiento y que los modelos mostraban tasas de falsos negativos más elevadas para las mujeres, lo que demuestra cómo los datos sesgados pueden contribuir a **un rendimiento diagnóstico desigual**.[7] Ampliando esta preocupación al ámbito de la equidad racial, otro estudio ha revelado que los modelos de IA entrenados para predecir la gravedad de la depresión a partir del lenguaje utilizado en las redes sociales funcionaban bien con participantes blancos, pero mal con participantes negros, lo que pone de manifiesto disparidades de rendimiento basadas en la raza que reflejan diferencias en los patrones de expresión que los modelos no captan.[8]

Abordar estos riesgos de sesgo requiere **modelos normativos que traduzcan los principios éticos y jurídicos internacionales y europeos en mecanismos de gobernanza concretos** que las instituciones sanitarias puedan aplicar en la práctica. En la Unión Europea (UE), el panorama normativo ha evolucionado rápidamente con la adopción de la Ley de Inteligencia Artificial (Ley de IA), que establece obligaciones para los sistemas de IA de alto riesgo y para los responsables de su implantación, como los hospitales.[9] Sin embargo, la aplicación de estos requisitos legales como procedimientos prácticos en las instituciones sanitarias sigue siendo un reto importante.

Este documento de políticas propone **un modelo regulador para la gobernanza de la IA biomédica que se ajusta al acervo de la UE**, incluida la Carta de los Derechos Fundamentales de la UE, y está diseñado para los sistemas sanitarios públicos y las organizaciones de la sociedad civil (OSC). El modelo propuesto se basa en instrumentos normativos europeos, entre los que se incluyen la Ley de IA, el Reglamento sobre productos sanitarios (MDR), el Reglamento sobre productos sanitarios para diagnóstico in vitro (IVDR), el Reglamento General de Protección de Datos (RGPD) y el emergente Espacio Europeo de Datos Sanitarios (EHDS). Este modelo traduce los principios de los

---

[6] Sorkin, D.H., Ngo-Metzger, Q., Billimek, J., August, K.J., Greenfield, S., & Kaplan, S.H., Underdiagnosed and Undertreated Depression Among Racially/Ethnically Diverse Patients With Type 2 Diabetes, *Diabetes Care* 34(3): 598–600, 2011.

[7] Straw, I., Rees, G., & Nachev, P., *Sex-Based Performance Disparities in Machine Learning Algorithms for Cardiac Disease Prediction: Exploratory Study*, *Journal of Medical Internet Research* 26: e46936, 2024.

[8] Rai, S., Stade, E. C., Giorgi, S., Francisco, A., Ungar, L. H., Curtis, B., & Guntuku, S. C., *Key language markers of depression on social media depend on race*, *Proceedings of the National Academy of Sciences* 121(14): e2319837121, 2024.

[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council, *Laying Down Harmonised Rules on Artificial Intelligence (AI Act)*, PE/24/2024/REV/1, Official Journal of the European Union, 13 June 2024.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

derechos fundamentales —en particular, la dignidad humana, la igualdad, la no discriminación y el derecho a la asistencia sanitaria— en garantías prácticas aplicables a todo el ciclo de vida de los sistemas de IA. Las referencias a ejemplos seleccionados de jurisdicciones no pertenecientes a la UE se incluyen únicamente cuando aportan conocimientos transferibles y tienen por objeto complementar el marco jurídico europeo basado en los derechos que sustenta el modelo.

Además, el modelo regulador se basa en enfoques normativos consolidados de otros sectores de alto riesgo, como la gestión de la seguridad aérea, la gobernanza del riesgo de los modelos financieros, la farmacovigilancia en la regulación de los medicamentos y las evaluaciones de impacto algorítmico en la administración pública. Estos enfoques aportan valiosas lecciones para la creación de marcos de gobernanza capaces de abordar los complejos riesgos del ciclo de vida asociados a la IA biomédica y se han adaptado aquí para integrar explícitamente consideraciones de equidad de género y racial a lo largo de todo el ciclo de vida de la IA.

El modelo proporciona a las instituciones sanitarias y a las OSC un **marco de gobernanza práctico que abarca la adquisición, la validación, la implantación, el seguimiento y la rendición de cuentas** de los sistemas de IA biomédica. Al integrar la protección de los derechos fundamentales en los procesos operativos, este marco tiene por objeto reducir los riesgos de sesgo de género y racial en la toma de decisiones sanitarias respaldadas por la IA y contribuir a unos resultados sanitarios más equitativos en toda la UE. Además, el documento **ofrece un conjunto de políticas, herramientas y recursos prácticos que los hospitales y las OSC pueden adaptar a los distintos contextos nacionales** para aplicar el modelo de manera coherente en entornos reales.

El marco jurídico europeo para la IA en el ámbito sanitario

La UE ha establecido un entorno normativo de múltiples niveles para la IA biomédica, compuesto por cuatro niveles jurídicos complementarios: la Ley de IA, la legislación sobre productos sanitarios y las directrices relacionadas, la protección de datos y la gobernanza de los datos sanitarios.

La Ley de IA establece un marco regulador basado en el riesgo para los sistemas de IA. Muchas aplicaciones sanitarias, incluidas las herramientas de apoyo a la toma de decisiones clínicas y los algoritmos de diagnóstico, entran en la categoría de sistemas de IA de alto riesgo, que deben cumplir requisitos estrictos en materia de gestión de riesgos, gobernanza de datos, transparencia, supervisión humana y precisión. Los requisitos aplicables incluyen también la identificación y mitigación de los riesgos de resultados discriminatorios mediante un análisis desagregado de las poblaciones, considerando el sexo y el género, el origen étnico y otras categorías sociales como variables transversales. **Las instituciones sanitarias que implementan sistemas de IA**

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

también tienen obligaciones específicas en virtud de la Ley de IA. Entre ellas se incluyen garantizar que los sistemas se utilicen de acuerdo con las instrucciones del proveedor, supervisar su rendimiento, mantener registros operativos y notificar los incidentes graves a las autoridades pertinentes, entre otros casos, cuando dichos incidentes puedan afectar de manera desproporcionada a diferentes poblaciones de pacientes.

En lo que respecta a la aplicación de la Ley de IA, en 2025 ya entraron en vigor las prácticas prohibidas y también comenzaron a aplicarse algunas normas de supervisión y gobernanza.[10] El próximo hito importante es agosto de 2026, cuando entrarán en vigor la mayoría de las obligaciones previstas en la Ley (incluidos los sistemas de alto riesgo del anexo III y las obligaciones de transparencia del artículo 50). Está previsto que la IA de alto riesgo integrada en productos regulados (lo que incluye muchos sistemas de IA para dispositivos médicos) entre en vigor a partir de agosto de 2027.

Para los hospitales, las disposiciones de la Ley de IA más importantes desde el punto de vista práctico son:

- **Requisitos** para **los sistemas de alto riesgo**. Los proveedores deben implementar un sistema de gestión de riesgos a lo largo del ciclo de vida (art. 9) y prácticas de gobernanza de datos que incluyan: expectativas explícitas de detección y mitigación de sesgos (art. 10), incluida la evaluación de la representatividad de los datos y las posibles disparidades de rendimiento entre las poblaciones mediante un análisis desagregado por sexo y género y por origen racial o étnico; garantías de transparencia y suministro de información a los responsables del despliegue (art. 13); supervisión humana diseñada para prevenir el sesgo de la automatización y permitir procedimientos de anulación o interrupción (art. 14); y controles de precisión, solidez y ciberseguridad, prestando especial atención a los bucles de retroalimentación y a los ataques específicos contra la IA (art. 15), que, de no abordarse, podrían exacerbar las desigualdades existentes.

- **Obligaciones de los responsables de la implementación**. Los hospitales y otros responsables de la implementación deben utilizar los sistemas de IA de alto riesgo de conformidad con las instrucciones del proveedor (art. 26); asignar una supervisión humana competente (art. 26); garantizar que los datos de entrada sean representativos y completos cuando el responsable de la implementación controle dichas entradas (art. 26), así como (cuando sea factible) tener en cuenta la diversidad demográfica de las poblaciones clínicas; supervisar el funcionamiento del sistema (art. 26); conservar los registros generados automáticamente durante al menos seis meses (art. 26); e informar oportunamente de los incidentes graves a los proveedores y a las autoridades

[10] European Commission, *Timeline for Implementation of the EU AI Act*, Brussels: European Commission, 2024.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

competentes (arts. 26 y 73), especialmente cuando exista riesgo de resultados clínicos desiguales.

- **Evaluación del impacto sobre los derechos fundamentales (FRIA)**. Antes de implantar determinados sistemas de IA de alto riesgo, los organismos de derecho público (y las entidades privadas que prestan servicios públicos) deben realizar una FRIA, en la que se describan los procesos, los grupos afectados, los riesgos, las medidas de supervisión humana y los mecanismos de mitigación o de reclamación, y notificarla posteriormente a la autoridad de vigilancia del mercado (art. 28). Este requisito es de relevancia directa para los hospitales públicos de muchos Estados miembros, ya que proporciona un mecanismo para identificar y abordar, antes de la implantación, los riesgos de discriminación y los efectos desiguales evaluados mediante un análisis desglosado por sexo y género, así como por raza o etnia.

Más allá de la Ley de IA, **los sistemas de IA utilizados en contextos médicos también están sujetos a la legislación sobre productos sanitarios y a las directrices relacionadas**, entre ellas el *Reglamento sobre productos sanitarios (MDR)*[11] y el *Reglamento sobre productos sanitarios para diagnóstico in vitro (IVDR),*[12] , que regulan la seguridad y el rendimiento de las tecnologías médicas. Estos reglamentos exigen una evaluación clínica y una vigilancia posterior a la comercialización del software de los productos sanitarios, incluidas las herramientas basadas en la IA, prestando atención a un rendimiento equitativo en las poblaciones de pacientes desglosadas, siempre que los datos lo permitan. Además, la Guía del Grupo de Coordinación de Productos Sanitarios (MDCG) *sobre la cualificación y clasificación del software en el marco del MDR/IVDR* ofrece directrices sobre cómo cualificar y clasificar el software con arreglo al MDR y al IVDR.[13] En ella se hace hincapié en que la clasificación depende de la finalidad prevista y no de la ubicación del software (por ejemplo, basado en la nube o integrado), e incluye ejemplos extraídos de sistemas de información hospitalaria y software de apoyo a la toma de decisiones clínicas, destacando la necesidad de tener en cuenta los impactos diferenciales en las diversas poblaciones de pacientes. *La Guía* del MDCG *sobre la evaluación clínica y de rendimiento del software de productos sanitarios* ofrece apoyo adicional, reforzando la expectativa de que la evidencia debe ser proporcional a la función clínica del software, incluida la evidencia de que el rendimiento de la IA es

[11] European Union, *Regulation (EU) 2017/745 on Medical Devices (MDR)*, Official Journal of the European Union, L117, Brussels: EU, 2017.

[12] European Union, *Regulation (EU) 2017/746 on In Vitro Diagnostic Medical Devices (IVDR)*, Official Journal of the European Union, L117, Brussels: EU, 2017.

[13] Medical Device Coordination Group, *Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 (MDR) and Regulation (EU) 2017/746 (IVDR)*, Brussels: MDCG, 2019.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

coherente en todas las poblaciones diferenciadas por sexo y género, así como por origen racial o étnico.[14]

La relación entre la Ley de IA y la legislación vigente sobre productos sanitarios se aclara en la *«Guía sobre la interacción entre la Ley de IA y el MDR/IVDR»*.[15] El documento explica que la Ley de IA se aplica junto con estos dos reglamentos. En la práctica, esto significa **que** muchos **sistemas de IA médica deben cumplir con ambos marcos normativos simultáneamente**. La guía también aclara cuándo la IA médica se considera de alto riesgo en virtud de la Ley de IA, lo que suele ocurrir cuando el propio sistema de IA es un producto sanitario (o un componente de seguridad del mismo) y, por lo tanto, está sujeto a una evaluación de conformidad por parte de un organismo notificado. Esto tiene implicaciones prácticas para la contratación pública hospitalaria. Los hospitales deberían exigir a los proveedores que demuestren la situación normativa de sus sistemas, incluida la clasificación según el MDR/IVDR, la intervención de un organismo notificado cuando proceda y el cumplimiento de los requisitos de alto riesgo de la Ley de IA, así como la documentación sobre el rendimiento desglosado por grupos de población y las medidas de mitigación del sesgo cuando sea pertinente. Los contratos de adquisición también deberían exigir a los proveedores que faciliten la documentación, la información sobre transparencia y las medidas de supervisión que necesitan los responsables de la implantación en virtud de la Ley de IA, incluidas disposiciones para garantizar resultados equitativos para todos los pacientes.

Varios **instrumentos adicionales de la UE también son relevantes para la gobernanza de la IA en el ámbito sanitario en la práctica**. El *Reglamento General de Protección de Datos (RGPD)* sigue siendo una limitación fundamental para la IA biomédica, ya que los datos sanitarios son datos personales de categorías especiales y las implementaciones hospitalarias suelen requerir evaluaciones de impacto en materia de protección de datos (EIPD). La Ley de IA (artículos 26, apartado 9, y 27, apartado 4) vincula explícitamente la práctica de los responsables de la implementación con las obligaciones derivadas de la EIPD (uso de la información del proveedor para respaldar la EIPD) y sitúa a la FRIA como complementaria a la EIPD, lo que puede incluir la consideración de los impactos dispares en las poblaciones diferenciadas por sexo y género y por origen racial o étnico. Además, el *Espacio Europeo de Datos Sanitarios (EHDS)* tiene por objeto facilitar el acceso seguro a los datos sanitarios para la prestación de asistencia sanitaria y la investigación, al tiempo que mantiene sólidas garantías de protección de datos.[16] Se espera que el EHDS desempeñe un papel clave a la hora de facilitar conjuntos de datos

[14] Medical Device Coordination Group, *Guidance on Clinical Evaluation (MDR) and Performance Evaluation (IVDR) of Software in Medical Devices*, Brussels: MDCG, 2020.

[15] Medical Device Coordination Group & European AI Board, *Guidance on MDR/IVDR and AI Act Interplay*, Brussels: European Commission, 2025.

[16] European Union, *Regulation on the European Health Data Space (EHDS)*, Brussels: European Union, 2024.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

más representativos para el desarrollo de la IA biomédica, respaldando una evaluación centrada en la equidad y reduciendo el riesgo de infrarrepresentación de poblaciones de pacientes con determinadas características demográficas.[17]

Por último, en lo que respecta a la ciberseguridad, la Agencia de Ciberseguridad de la Unión Europea ha elaborado unas directrices sobre prácticas de ciberseguridad para la IA. En ellas se sostiene que las normas organizativas existentes (por ejemplo, las normas ISO de gestión de la seguridad y de gestión de la calidad) pueden adaptarse a los contextos de la IA.[18] La *Guía sobre ciberseguridad para productos sanitarios* del MDCG describe cómo deben cumplir los fabricantes los requisitos relacionados con la ciberseguridad del MDR y el IVDR mediante la gestión de riesgos y unas expectativas mínimas de ciberseguridad.[19] Es importante destacar que una sólida gobernanza en materia de ciberseguridad sustenta un despliegue equitativo de la IA: proteger la integridad de los datos, el acceso a los mismos y la fiabilidad de los sistemas ayuda a evitar la amplificación involuntaria de las desigualdades de género y raciales o étnicas.

En conjunto, estos instrumentos jurídicos crean un entorno normativo complejo en el que deben operar los sistemas de IA aplicados a la asistencia sanitaria. El modelo regulador de la IA para los hospitales públicos europeos que aquí se propone se ha diseñado explícitamente en torno a esas cuatro «capas» jurídicas fundamentales que rigen la práctica de la IA en el ámbito sanitario. Respectivamente, el modelo prevé mecanismos para supervisar y mitigar los impactos desiguales en las poblaciones diferenciadas por sexo y género y por origen racial o étnico a lo largo de todo el ciclo de vida de la IA.

Marcos de gobernanza europeos e internacionales aplicables a una IA responsable en el ámbito sanitario

Es importante señalar que los marcos normativos europeos mencionados anteriormente constituyen la base jurídica principal para la gobernanza de la IA biomédica en la Unión. Además del marco normativo de la UE en materia de IA biomédica, existen varios instrumentos internacionales no vinculantes que ofrecen orientaciones para la gestión de los riesgos de la IA. En particular, los **marcos de gobernanza adicionales que se repasan en esta sección se incluyen para mostrar técnicas de aplicación que pueden trasladarse al contexto de la UE, sin dejar de estar**

[17] European Commission, *European Health Data Space (EHDS)*, Brussels: European Commission, 2025.

[18] European Union Agency for Cybersecurity (ENISA), *Cybersecurity of AI and Standardisation*, Athens: ENISA, 2023.

[19] Medical Device Coordination Group, *MDCG 2019-16 Rev. 1 – Guidance on Cybersecurity for Medical Devices*, Brussels: European Commission, 2020.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

plenamente subordinadas a los requisitos jurídicos europeos y a los principios de los derechos fundamentales. Se utilizan de forma selectiva como ejemplos de otras jurisdicciones en las que se han creado en la práctica mecanismos comparables de gobernanza del ciclo de vida, especialmente en ámbitos como la gestión del riesgo de los modelos, la auditoría algorítmica y la evaluación del impacto institucional.

Por ejemplo, *el Marco de Gestión de Riesgos de la IA* (RMF) del Instituto Nacional de Estándares y Tecnología (NIST) ofrece un enfoque integral para la gestión de los riesgos de la IA a través de un enfoque basado en el ciclo de vida organizado en torno a la gobernanza, la identificación de riesgos, la medición de riesgos y la gestión de riesgos.[20] El marco reconoce explícitamente la equidad y la lucha contra los sesgos perjudiciales como dimensiones fundamentales de una IA fiable. Además, como parte del mapeo y la medición de riesgos, anima a las organizaciones a identificar y evaluar los impactos dispares en las distintas poblaciones de pacientes, teniendo en cuenta también las diferencias de sexo y género, así como las raciales y étnicas. El objetivo es integrar la fiabilidad y los controles de riesgo a lo largo del diseño, la implementación y la supervisión, combinando la evaluación técnica con la rendición de cuentas organizativa. Para los hospitales, el punto fuerte del Marco de Gestión de Riesgos de la IA (AI RMF) radica en su capacidad para traducirse en prácticas concretas, como **funciones definidas para la supervisión de la equidad, la evaluación metódica del rendimiento desglosado por poblaciones y estrategias de mitigación documentadas**, al tiempo que se mantiene una perspectiva más amplia basada en los derechos.

En lo que respecta al sistema de gestión, la norma *ISO/IEC 42001* de la Organización Internacional de Normalización establece los requisitos para un sistema organizativo de gestión de la IA, centrándose en las políticas, los objetivos, los procesos y la mejora continua para una IA responsable.[21] Aunque no es prescriptiva en cuanto a daños específicos, apoya la integración de los principios de equidad y no discriminación en las políticas organizativas y los controles de riesgo, lo que permite a las instituciones abordar los posibles sesgos de género y raciales de forma coordinada. Como complemento a ello, *la norma ISO/IEC 23894* ofrece orientación sobre la integración de la gestión de riesgos en el desarrollo y la implantación de la IA, incluida la identificación y el tratamiento de los riesgos relacionados con impactos no deseados o desproporcionados en diferentes grupos de población.[22] En lo que respecta específicamente a la IA clínica, *la norma ISO 14971* define los principios de gestión de riesgos ampliamente utilizados en la regulación de los productos sanitarios, que

[20] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, Gaithersburg, MD: NIST, 2023.

[21] International Organization for Standardization, *ISO/IEC 42001:2023 – Artificial Intelligence Management System*, Geneva: ISO, 2023.

[22] International Organization for Standardization, *ISO/IEC 23894:2023 – Artificial Intelligence Risk Management*, Geneva: ISO, 2023.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

describen la identificación sistemática de peligros, la evaluación de riesgos, el control de riesgos y la supervisión del ciclo de vida.[23] Estos procesos pueden ampliarse para tener en cuenta el rendimiento o los resultados desiguales entre poblaciones desglosadas como parte de la identificación de daños. Las normas descritas son valiosas, ya que sugieren **una estructura de sistema de gestión que los hospitales pueden adoptar**. También proporcionan un lenguaje común a través del cual se pueden especificar los requisitos relacionados con la equidad —como la evaluación de poblaciones desglosadas y la mitigación de sesgos— en la contratación pública y en los acuerdos con los proveedores.

Los marcos de gobernanza ética también proporcionan una importante orientación normativa. *Las Directrices éticas* de la CE *para una IA fiable* son un instrumento político útil, aunque no vinculante.[24] Definen siete requisitos fundamentales para los sistemas de IA responsables, entre los que se incluyen la diversidad, la no discriminación y la equidad, que exigen explícitamente evitar los sesgos y tener en cuenta los impactos en poblaciones con diferentes características demográficas, como el sexo y el género y el origen racial o étnico. *La Lista de Evaluación para una Inteligencia Artificial Fiable (ALTAI)*, una herramienta voluntaria de autoevaluación desarrollada por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la CE, aplica dichas directrices en forma de lista de autoevaluación.[25] En un modelo regulador sanitario, el valor de ALTAI reside en su **formato basado en preguntas, que puede adaptarse a listas de verificación prácticas para la adquisición y la implantación**, lo que anima a las partes interesadas a tener en cuenta cuestiones como la representatividad de los datos de entrenamiento, el riesgo de resultados clínicos dispares y la idoneidad de las salvaguardias contra los efectos discriminatorios, incorporando así las consideraciones de equidad en la toma de decisiones rutinaria por parte de los profesionales clínicos, los especialistas en TI y los gestores hospitalarios.

Desde una perspectiva de derechos humanos, el *Convenio Marco* del Consejo de Europa *sobre la IA* es un instrumento jurídico relevante que se basa explícitamente en los principios de dignidad humana, transparencia, rendición de cuentas, igualdad y no discriminación, y privacidad, proporcionando así una base jurídica para abordar los efectos discriminatorios, incluidos los relacionados con los sesgos de género y raciales en los sistemas de IA.[26] Del mismo modo, *los Principios sobre la Inteligencia Artificial* de

---

[23] International Organization for Standardization, *ISO 14971:2019 – Medical Devices: Application of Risk Management to Medical Devices*, Geneva: ISO, 2019.

[24] European Commission, *Ethics Guidelines for Trustworthy AI*, Brussels: European Commission, 2019.

[25] European Commission, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) – Self-Assessment*, Brussels: European Commission, 2020.

[26] Council of Europe, *Framework Convention on Artificial Intelligence (CAI)*, Strasbourg: Council of Europe, 2023.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

la Organización para la Cooperación y el Desarrollo Económicos[27] y *la Recomendación* de la UNESCO *sobre la ética de la IA*[28] ofrecen orientaciones normativas internacionales que fomentan la identificación y la mitigación de los impactos injustos o dispares en las distintas poblaciones, reforzando la expectativa de que los sistemas de IA no deben reproducir ni amplificar las desigualdades estructurales. En lo que respecta a la gobernanza ética específica en el ámbito de la salud, la Organización Mundial de la Salud ha articulado seis principios fundamentales para la IA en la salud: autonomía, bienestar, explicabilidad, responsabilidad, equidad y capacidad de respuesta.[29] El principio de **equidad exige explícitamente que se preste atención a las diferencias en el acceso, el rendimiento y los resultados entre las diversas poblaciones de pacientes, teniendo en cuenta también el sexo y el género, así como las diferencias raciales o étnicas**.

Estos marcos de gobernanza internacionales se corresponden directamente con los requisitos de la Carta en materia de dignidad, igualdad, no discriminación y derecho a la salud, y comparten un énfasis común en la gobernanza a lo largo del ciclo de vida — principios que pueden traducirse directamente en controles prácticos para las instituciones sanitarias.

Lecciones de los modelos reguladores intersectoriales

Sectores de alto riesgo como la aviación, la banca y la industria farmacéutica llevan mucho tiempo desarrollando sofisticados sistemas de gobernanza para gestionar los riesgos tecnológicos. Estos sectores ofrecen valiosos enfoques normativos que pueden adaptarse a la gobernanza de la IA en el ámbito sanitario y a la gestión de los retos relacionados con los sesgos de género y raciales.

Aunque los sesgos de género y raciales no son riesgos tecnológicos en sentido estricto, se convierten en riesgos sociotécnicos cuando los datos, los algoritmos y las prácticas organizativas interactúan con las desigualdades estructurales en la asistencia sanitaria. En este contexto, el término «riesgo» se refiere a la manifestación a nivel del sistema de los sesgos en los resultados, las decisiones y los desenlaces clínicos, más que a su origen. La propia asistencia sanitaria ya ofrece amplias pruebas de tales efectos, incluidas las disparidades documentadas en el diagnóstico, el tratamiento y la predicción de riesgos entre poblaciones diferenciadas por sexo, género y raza o etnia. Por lo tanto, los ejemplos intersectoriales complementan las pruebas de la asistencia sanitaria al ilustrar mecanismos de gobernanza consolidados para gestionar sistemas complejos y de gran

[27] Organisation for Economic Co-operation and Development, *OECD Principles on Artificial Intelligence*, Paris: OECD, 2019.

[28] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, Paris: UNESCO, 2021.

[29] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health*, Geneva: WHO, 2021.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

impacto, junto con los hallazgos empíricos de la asistencia sanitaria que demuestran la necesidad de dichos controles normativos.

Un ejemplo especialmente relevante es el *Sistema de Gestión de la Seguridad (SMS)* utilizado en la aviación. La normativa internacional de aviación civil no considera la seguridad como un ejercicio de certificación puntual, sino como un proceso continuo en toda la organización que permite tomar decisiones basadas en la identificación de peligros, la gestión de riesgos, la garantía de la seguridad (seguimiento y auditoría), la gestión del cambio y la mejora continua.[30] Aplicar esta lógica a la IA en el ámbito sanitario significaría **establecer un «Sistema de Gestión de la Seguridad de la IA» que incorpore explícitamente consideraciones de equidad**, entre las que se incluyen: (1) un canal de notificación de peligros y cuasiaccidentes para que el personal clínico registre posibles incidentes de sesgo de género o racial; (2) un registro formal de riesgos para cada sistema de IA que documente las diferencias de rendimiento entre poblaciones de pacientes desglosadas; (3) mecanismos de control de cambios que evalúen el impacto de las actualizaciones de los modelos, la deriva de los datos o los cambios en los flujos de trabajo sobre los resultados en materia de equidad; y (4) revisiones periódicas de garantía de seguridad con medidas correctivas que aborden tanto los riesgos técnicos como las desigualdades de género o raciales identificadas. El énfasis explícito de la Ley de IA (artículos 9 y 72) en la gestión de riesgos a lo largo del ciclo de vida y la supervisión posterior a la comercialización se ajusta estrechamente a esta lógica del SMS.

Un segundo enfoque útil proviene de la regulación financiera, donde las herramientas cuantitativas complejas se tratan como «modelos» que requieren un desarrollo sólido, una validación independiente y una supervisión de gobernanza. En la supervisión bancaria, se han implementado marcos de gobernanza de riesgos comparables en jurisdicciones como Estados Unidos (por ejemplo, *la Guía de Supervisión sobre la Gestión del Riesgo de Modelos* de la Reserva Federal de EE. UU.*)* y ofrecen un ejemplo de cómo se pueden implementar la validación, la supervisión y la revisión independiente de los modelos.[31] Se podrían aplicar enfoques similares a los sistemas de IA biomédica para garantizar que **las herramientas de apoyo a la toma de decisiones clínicas se validen de forma independiente (interna o externamente) antes de su introducción** en entornos sanitarios, con **una evaluación explícita del rendimiento en poblaciones diferenciadas por sexo y género, y por características raciales o étnicas**. Los elementos mínimos de validación deben documentar la finalidad prevista del sistema, sus límites de funcionamiento, los modos de fallo conocidos y el rendimiento desglosado por poblaciones, incluidas las posibles disparidades o las medidas de mitigación aplicadas.

[30] International Civil Aviation Organization, *Safety Management Manual (Doc 9859),* 4th ed., Montréal: ICAO, 2018.

[31] Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency, *Supervisory Guidance on Model Risk Management (SR 11-7),* Washington, D.C., 4 April 2011.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

La revalidación debe activarse siempre que el sistema se implemente en un nuevo centro, se aplique a una población de pacientes diferente, se conecte a nuevos equipos o protocolos, sea actualizado por el proveedor o muestre signos de desviación en su rendimiento. Estas prácticas son coherentes con la Ley de IA (artículos 13, 14 y 60), que exige la realización de pruebas, la transparencia para los responsables de la implementación y la supervisión humana destinada a detectar y prevenir el sesgo de la automatización, garantizando que la atención asistida por IA sea segura y equitativa en todas las poblaciones diversas.

En lo que respecta a la gobernanza tras la implementación, la regulación farmacéutica ofrece un modelo sólido: en farmacovigilancia, las reacciones adversas y las señales de seguridad se recogen y analizan sistemáticamente para detectar riesgos emergentes. En concreto, *el Módulo VI de las Buenas Prácticas de Farmacovigilancia (GVP)* de la Agencia Europea de Medicamentos establece requisitos detallados para la recogida, gestión y presentación de informes sobre sospechas de reacciones adversas, lo que refleja un enfoque ecosistémico para la notificación y el seguimiento de incidentes.[32] **Un sistema** comparable **de «vigilancia de la IA» podría permitir a los hospitales notificar y analizar incidentes** relacionados con decisiones clínicas respaldadas por la IA, prestando especial atención a los resultados en materia de equidad. Entre los elementos clave se incluirían: (1) la notificación sistemática que recoja cualquier desigualdad observada en la atención o en los resultados en función del sexo y el género o de las diferencias raciales o étnicas; (2) la documentación obligatoria de los incidentes graves en los que el sesgo pueda haber contribuido a un trato diferencial; (3) la obligación de notificar al prestador y a las autoridades pertinentes cualquier preocupación relacionada con la equidad (notificando a las autoridades si el sesgo provoca un perjuicio grave, como un diagnóstico erróneo o un trato desigual, mientras que las desigualdades rutinarias se mantendrían dentro del control de calidad interno); y (4) funciones de detección de tendencias y señales que identifiquen patrones emergentes de sesgo en los distintos departamentos, centros o en la red sanitaria en general. Dicho seguimiento permitiría adoptar medidas de mitigación proactivas y llevar a cabo una mejora continua, garantizando que la atención asistida por IA no solo sea segura, sino también equitativa.

Por último, las herramientas de gobernanza del sector público, como las evaluaciones de impacto algorítmico y las auditorías de sesgos, ofrecen métodos prácticos para evaluar el impacto social de los sistemas de toma de decisiones automatizados, incluidas las posibles disparidades entre sexos y géneros o las diferencias raciales y étnicas. Estas herramientas traducen los principios éticos y las normas de derechos humanos en formatos aplicables que las organizaciones pueden utilizar para anticipar y mitigar las desigualdades antes de la implantación. Por ejemplo, *la Evaluación de Impacto Algorítmico* federal de Canadá califica los riesgos de los sistemas de IA y los vincula a

[32] European Medicines Agency, *GVP Module VI: Collection and Submission of Suspected Adverse Reactions*, London: EMA, 2017.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

medidas de mitigación.[33] Esta herramienta enmarca explícitamente la implantación como un problema de gobernanza impulsado por el impacto. Además, el enfoque de la ciudad de Nueva York respecto a las herramientas automatizadas de toma de decisiones en materia de empleo exige auditorías de sesgos y obligaciones de divulgación ante el público.[34] En consecuencia, los hospitales y las organizaciones de la sociedad civil que trabajan en el ámbito sanitario podrían implementar un **paquete de «Evaluación y auditoría del impacto de la IA» que combine un cuestionario estructurado centrado en la equidad con auditorías periódicas e independientes sobre la equidad**, generando un resumen publicable del rendimiento en las distintas poblaciones desglosadas y de las medidas de mitigación. La implementación de dicho paquete contribuiría a promover la transparencia, la rendición de cuentas pública y la confianza de los y las pacientes en la atención asistida por IA.

La inclusión de modelos normativos de sectores como la seguridad aérea, la gestión del riesgo de los modelos financieros, la farmacovigilancia y la gobernanza algorítmica pretende constituir una transferencia deliberada de principios de gobernanza consolidados. Estos ámbitos se han seleccionado porque aplican de forma práctica enfoques formalizados para la detección, el seguimiento y la mitigación de riesgos en sistemas complejos en los que pueden producirse impactos diferenciales sobre las personas. Varios de ellos ya incorporan prácticas de equidad en la práctica, como la estratificación demográfica del riesgo en la farmacovigilancia, la validación de modelos y los controles de sesgo en la supervisión financiera, y la notificación sistemática de eventos adversos que afectan a poblaciones específicas. Su relevancia radica tanto en la similitud conceptual como en su uso contrastado de la gobernanza a lo largo del ciclo de vida, la notificación obligatoria y los mecanismos de supervisión independientes, que pueden adaptarse directamente para abordar los riesgos de equidad de género y racial en los sistemas de IA biomédica.

Plan para un modelo regulador de la IA alineado con la Carta de los Derechos Fundamentales de la UE para hospitales públicos y organizaciones de la sociedad civil

En esta sección se expone una arquitectura de modelo regulador que se **basa en los principios de la Carta de la UE, pero que se plasma en procedimientos prácticos que los hospitales y las OSC pueden aplicar** en la práctica. La Carta proporciona la base normativa para la gobernanza de las tecnologías sanitarias en Europa, en particular a

[33] Treasury Board of Canada Secretariat, *Algorithmic Impact Assessment*, Ottawa: Government of Canada, 2026.

[34] New York City Department of Consumer and Worker Protection, *Automated Employment Decision Tools*, New York: City of New York, 2023.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

través de los principios de dignidad humana, igualdad, no discriminación y derecho a la asistencia sanitaria.[35] Estos derechos son especialmente relevantes cuando los sistemas automatizados influyen en decisiones clínicas que afectan a la atención e a de los pacientes, incluidas las posibles disparidades entre poblaciones diferenciadas por sexo y género y origen racial o étnico.

En el nivel más fundamental, el artículo 1 de la Carta, que establece que la dignidad humana es inviolable, respalda los requisitos de gobernanza que garantizan que las herramientas de IA se utilicen de manera que se preserve el juicio humano y se respete la autonomía del paciente. En la práctica, este principio exige **salvaguardias contra la dependencia excesiva de las recomendaciones automatizadas, mecanismos para impugnar las decisiones asistidas por IA y estructuras de supervisión clínica que eviten el sesgo de la automatización** —la tendencia de los responsables humanos de la toma de decisiones a confiar excesivamente en los resultados algorítmicos, lo que puede amplificar las desigualdades si las recomendaciones de la IA perjudican sistemáticamente a determinadas poblaciones desagregadas—.

Del mismo modo, los artículos 20 y 21, que consagran la igualdad ante la ley y prohíben la discriminación, proporcionan la base jurídica para obligaciones explícitas de equidad y mitigación de sesgos en la IA biomédica. Por lo tanto, los hospitales que implementen sistemas de IA deberían poner en marcha **procesos para supervisar el rendimiento del sistema en diversas poblaciones de pacientes**, incluyendo el sexo y el género, la raza o el origen étnico, y las combinaciones interseccionales de estas características. Cuando se detecten desigualdades en el rendimiento o en los resultados, deberían aplicarse medidas correctivas mediante el ajuste de los modelos, cambios en el flujo de trabajo clínico o vías de decisión alternativas, garantizando que la implementación de la IA no consolide las desigualdades estructurales.

Es más, el artículo 35 de la Carta garantiza el derecho a un alto nivel de protección de la salud humana y, por lo tanto, respalda aún más el principio de que los sistemas de IA solo pueden implantarse cuando contribuyan de forma demostrable a una asistencia sanitaria segura, eficaz y equitativa. Las tecnologías de IA no deben introducir nuevas barreras a la atención sanitaria ni exacerbar las disparidades existentes que afectan a poblaciones de pacientes históricamente desatendidas o en situación de desventaja, y su impacto en dichas poblaciones debe evaluarse de forma continua.

Para garantizar estos derechos en la práctica, el modelo regulador propuesto en este documento de política adopta un **enfoque de gobernanza basado en el ciclo de vida que estructura la supervisión de la IA en torno a una serie de puntos de decisión obligatorios o «puerta de control»**. Estas «puertas de control» reflejan la lógica

[35] European Union, *Charter of Fundamental Rights of the European Union*, 2012/C 326/02, Strasbourg: EU.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

reguladora incorporada en la Ley de IA de la UE y en la legislación sanitaria relacionada, al tiempo que proporcionan a los hospitales una vía de gobernanza clara y auditable. El modelo parte de la base de que el hospital suele ser un «implementador» (en virtud de la Ley de IA) y un «prestador de asistencia sanitaria» según la legislación sanitaria nacional, pero también aborda el caso en el que el hospital se convierte en «prestador/fabricante» (como en el ejemplo de la IA interna).

Los hospitales son los principales implementadores y responsables de la implementación de los sistemas de IA, mientras que las organizaciones de la sociedad civil (OSC) actúan como organismos independientes de supervisión, auditoría y seguimiento basado en los derechos, sin control diario sobre las decisiones de implementación clínica.

La aplicación de este modelo regulador se rige por un principio de proporcionalidad que reconoce que **los hospitales y las OSC pueden presentar distintos niveles de capacidad técnica, organizativa y analítica en los distintos Estados miembros y sistemas sanitarios**. Se aplican requisitos mínimos obligatorios de manera uniforme para garantizar unas salvaguardias básicas en materia de seguridad, transparencia y equidad. Además, las prácticas avanzadas, como el análisis ampliado y desagregado de la población, la auditoría externa y las herramientas sofisticadas de control del sesgo, pueden implementarse de forma progresiva en función de la capacidad institucional. El anexo técnico también sigue este enfoque al distinguir entre requisitos mínimos y recomendados.

1. **Ámbito de aplicación y clasificación.** Antes de la adquisición o el desarrollo, las instituciones deben determinar la situación normativa de una herramienta de IA determinada. Esto incluye evaluar si el sistema cumple los requisitos para ser considerado software de dispositivo médico en virtud del MDR o del IVDR, si constituye un sistema de IA de alto riesgo según la Ley de IA, o si entra en una categoría de menor riesgo que, no obstante, requiere supervisión regulatoria. En esta fase inicial, los hospitales pueden señalar aquellas herramientas cuyo uso previsto pueda tener implicaciones más marcadas en materia de equidad o afectar a poblaciones de pacientes diversas, teniendo en cuenta que estas herramientas requerirán una evaluación más detallada de los sesgos de género y raciales en fases posteriores. Las directrices elaboradas por el MDCG sobre la cualificación y clasificación del software ofrecen ejemplos prácticos que pueden ayudar a evitar una clasificación normativa errónea y sentar las bases para las

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

evaluaciones de sesgos y equidad.[36] Los hospitales lideran la clasificación y la determinación normativa. Los responsables de la seguridad clínica (CSO) aportan asesoramiento independiente sobre los riesgos relacionados con la equidad, pero no tienen autoridad para tomar decisiones en esta fase.

2. **Fase previa a la contratación y a la firma del contrato.** Los hospitales suelen ejercer su mayor influencia sobre la gobernanza de la IA durante el proceso de contratación. En esta fase, las instituciones deben asegurarse de que los proveedores se comprometan contractualmente a abordar los riesgos de equidad, incluidos los sesgos de género y raciales, a lo largo de todo el ciclo de vida de la IA. Se debe exigir a los proveedores que faciliten documentación que demuestre el cumplimiento normativo, incluidas pruebas de evaluaciones de conformidad cuando sea pertinente, resultados de validación del rendimiento e información necesaria para que los implementadores cumplan con sus obligaciones en virtud de la Ley de IA. Los contratos de adquisición deben asignar explícitamente las responsabilidades entre el proveedor y el responsable de la implementación, incluidos los requisitos relativos a la transparencia, la documentación, el registro, la supervisión humana y el apoyo posterior a la implementación. El objetivo en esta etapa es integrar la concienciación sobre los sesgos y la rendición de cuentas en las obligaciones contractuales, sentando así las bases para una implementación segura y equitativa. Los hospitales son responsables de las decisiones de adquisición y de los acuerdos contractuales. Las organizaciones de la sociedad civil (OSC) pueden revisar la documentación de la adquisición en lo que respecta a consideraciones de equidad y derechos, y emitir recomendaciones no vinculantes.
3. **Etapa de validación.** La tercera etapa es una etapa de validación, que garantiza que las herramientas de IA se evalúen de forma independiente antes de incorporarse a los flujos de trabajo clínicos. La validación debe evaluar tanto el rendimiento general como los resultados en poblaciones de pacientes desglosadas, prestando especial atención al sexo y al género, así como al origen racial o étnico. Este paso es fundamental, ya que los sistemas de IA biomédica pueden mostrar un rendimiento desigual entre las distintas poblaciones cuando los datos de entrenamiento no son representativos o cuando se pasan por alto las diferencias contextuales entre los entornos de desarrollo y de implementación.[37] La evaluación también debe documentar cualquier medida de mitigación aplicada y poner de relieve los riesgos de equidad restantes, proporcionando a los hospitales pruebas que les permitan tomar decisiones de

[36] Medical Device Coordination Group, *Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 (MDR) and Regulation (EU) 2017/746 (IVDR)*, Brussels: MDCG, 2019.

[37] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. y Chin, M. H., *«Ensuring Fairness in Machine Learning to Advance Health Equity»*, *Annals of Internal Medicine*, 169(12), 866-872, 2018.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

implementación fundamentadas. Los hospitales son responsables de llevar a cabo o encargar la validación e interpretar los resultados para tomar decisiones de implementación. Los responsables de seguridad clínica (CSO) proporcionan una revisión independiente del rendimiento desglosado y de los riesgos de sesgo, sin tener control sobre los resultados de la validación.

4. **Control de implementación.** Una vez que un sistema ha superado la validación, la cuarta etapa es un control de implementación. Los hospitales deben garantizar que las condiciones de trabajo favorezcan un uso seguro, equitativo y responsable. Esto incluye asignar personal de supervisión debidamente formado, implementar mecanismos que permitan a profesionales clínicos anular o desactivar las recomendaciones algorítmicas, y asegurarse de que los usuarios comprendan la finalidad prevista del sistema, sus limitaciones y cualquier consideración desagregada sobre su rendimiento. Se debe informar a los y las profesionales clínicos de posibles variaciones injustas en los resultados cuando se tengan en cuenta diferencias de sexo y género o raciales/étnicas, y se debe informar a los y las pacientes cuando se utilicen herramientas de decisión asistidas por IA en su atención. Los hospitales asumen toda la responsabilidad sobre las decisiones de implementación, la integración clínica y la supervisión de la puesta en marcha. Las organizaciones de la sociedad civil (OSC) no intervienen en la implementación, pero pueden revisar la documentación sobre la preparación para la implementación desde una perspectiva de equidad y derechos.
5. **Etapa de seguimiento, auditoría y rendición de cuentas.** Esta última etapa rige la fase posterior al despliegue del ciclo de vida de la IA. Los sistemas de IA deben ser objeto de un seguimiento continuo en cuanto a su rendimiento general, las disparidades entre las poblaciones desglosadas y los incidentes relacionados con la seguridad o la equidad. Los hospitales deben mantener registros del funcionamiento del sistema, implementar procedimientos estandarizados de notificación de incidentes que detecten posibles sesgos de género o raciales, y reevaluar periódicamente el rendimiento del sistema mediante ejercicios de revalidación. Cuando surjan desigualdades o riesgos de sesgo que no puedan mitigarse de manera eficaz, el sistema deberá modificarse, suspenderse o retirarse del servicio. Las conclusiones deberán servir de base tanto para la gobernanza interna como, cuando proceda, para el aprendizaje a nivel de red más amplio, con el fin de prevenir sesgos recurrentes. Los hospitales son responsables del seguimiento continuo, el registro, la notificación de incidentes y la aplicación de medidas de mitigación. Las OSC llevan a cabo auditorías independientes, evalúan los riesgos sistémicos para la equidad en todos los centros y apoyan la escalación de las disparidades persistentes, sin tener control operativo sobre las modificaciones del sistema.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

En conjunto, estas etapas del ciclo de vida transforman los principios fundamentales de los derechos humanos en procesos de gobernanza concretos capaces de abordar los riesgos de sesgo de género y racial en los sistemas de IA clínica.

Figura 1: Vía de gobernanza de la IA alineada con la Carta de AEQUITAS

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Políticas prácticas, herramientas operativas y instrumentos de gobernanza para los profesionales de los hospitales y las OSC

Para funcionar de manera eficaz en entornos sanitarios reales, el modelo regulador propuesto en este documento de políticas prevé la creación de instrumentos prácticos que puedan reutilizarse en distintas instituciones y adaptarse a diferentes contextos nacionales. Hace hincapié en la necesidad de procedimientos, documentación y plantillas estandarizados que permitan tanto a los hospitales como a las OSC gestionar los sistemas de IA de manera coherente. El siguiente conjunto de instrumentos se ajusta a los requisitos de la Ley de IA, a las directrices relativas al MDR y al IVDR, y a las prácticas de gobernanza intersectoriales.

Fase de contratación pública y contratación

- **Paquete de contratación pública y contratación.** La comunidad de contratación pública en materia de IA de la UE ha publicado cláusulas contractuales tipo (versiones de alto riesgo y de bajo riesgo) y las ha actualizado para reflejar la Ley de IA, animando explícitamente a las autoridades públicas a utilizarlas y proporcionando comentarios para su aplicación.[38] En el ámbito de la IA en la asistencia sanitaria, con el fin de evaluar y mitigar los riesgos de sesgo de género y racial, estas cláusulas pueden adaptarse en un anexo específico para cada hospital, como documento independiente, que exija (como mínimo):
    - (1) la presentación de informes estandarizados sobre la composición de los datos de entrenamiento, incluida la representación en cuanto a sexo, género y origen racial o étnico, para respaldar la evaluación de la representatividad y el riesgo de sesgo;
    - (2) la divulgación de la procedencia de los datos, incluyendo si los conjuntos de datos son de propiedad exclusiva o no, sus condiciones de accesibilidad y las implicaciones para la transparencia, la validación independiente y la evaluación de los sesgos de género y raciales;
    - (3) la divulgación de los resultados en las distintas poblaciones desagregadas y de las pruebas de validación, incluidos los resultados estratificados por sexo y género y por origen racial o étnico, la documentación de cualquier disparidad conocida, las medidas de mitigación del sesgo aplicadas durante el desarrollo y las pruebas de validación que lo respalden (incluidas las consideraciones específicas del centro relevantes para la equidad y la imparcialidad);

[38] European Commission, *Updated EU AI Model Contractual Clauses Now Available*, Brussels: European Commission, 2025.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- (4) requisitos de registro y auditabilidad, que garanticen que el sistema es capaz de generar registros de uso que puedan servir de base para auditorías independientes sobre sesgos o impactos desiguales;
- (5) obligaciones de actualización y control de cambios, que exijan al proveedor informar al hospital de las actualizaciones del modelo, la recapacitación o los cambios en los flujos de trabajo, y documentar sus posibles efectos sobre el rendimiento con respecto a las poblaciones diferenciadas;
- (6) compromisos de respuesta ante incidentes relacionados con sesgos, que obliguen al proveedor a notificar los errores o desigualdades identificados que afecten a poblaciones de pacientes diferenciadas por sexo y género o por origen racial o étnico, y a esbozar las medidas correctivas;
- (7) una clara asignación de responsabilidades, en la que se especifiquen las obligaciones del proveedor y del responsable de la implementación en lo que respecta al cumplimiento de los requisitos de la Ley de IA y a la gestión de los riesgos relacionados con la equidad y los sesgos.

Fases de validación y previas a la implementación

- **Política de gobernanza y representatividad de los datos sanitarios.** Para mitigar los sesgos de género y raciales, todos los hospitales y organizaciones de la sociedad civil que implementen IA deberían adoptar una política formal que garantice que los datos utilizados con fines de entrenamiento, ajuste o implementación sean representativos, completos y precisos para las poblaciones atendidas. Esto incluye evaluar la composición demográfica, documentar las poblaciones de pacientes infrarrepresentadas y definir medidas de mitigación cuando sea necesario. La política debe abarcar tanto los conjuntos de datos internos como los de origen externo, incluidos los datos de propiedad exclusiva y de acceso abierto, y exigir la verificación del rendimiento desagregado y de cualquier estrategia de mitigación de sesgos aplicada por los proveedores. Dicha gobernanza se ajusta a la Ley de IA (art. 10), que exige el examen y la mitigación de los riesgos de sesgo, y refleja las mejores prácticas recogidas en la bibliografía sobre la mitigación de sesgos en la IA sanitaria, que abarca todo el ciclo de vida, desde la recopilación de datos hasta el seguimiento posterior a la implementación.[39] Esta política no es un artefacto en sí misma, sino una capa normativa que orienta a otros artefactos. Funciona como un principio previo a la implementación y se integra directamente en el seguimiento continuo, la auditoría y la rendición de cuentas.

[39] Char, D.S., Shah, N.H., & Magnus, D., *Implementing Machine Learning in Health Care — Addressing Ethical Challenges*, *New England Journal of Medicine* 378: 981-983 (2018).

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- **Referencia de ciberseguridad y modelo de amenazas específico para la IA.** Los sistemas de IA hospitalarios pueden ser vulnerables no solo a ataques convencionales, como el envenenamiento de datos, los ejemplos adversarios o la evasión de modelos, sino también a manipulaciones o errores que exacerben los sesgos de género o raciales, por ejemplo, al afectar de manera desproporcionada a los estratos demográficos infrarrepresentados en los datos de entrenamiento o implementación. La Ley de IA (art. 15) exige resiliencia frente a alteraciones no autorizadas de los resultados o el rendimiento, lo que pone de relieve la necesidad de salvaguardias específicas para la IA. Por lo tanto, los hospitales deben establecer una base de referencia de ciberseguridad y un modelo de amenazas específico para la IA que tenga en cuenta explícitamente los riesgos relacionados con los sesgos, además de las amenazas técnicas. Este documento en constante evolución funciona tanto como política como herramienta práctica, sirviendo de lista de comprobación previa a la implementación para garantizar la aplicación de controles de ciberseguridad alineados con el MDR y el IVDR, la gobernanza de la seguridad de la información de la organización (p. ej., ISO/IEC 27001) y los planes de pruebas y modelización de amenazas sensibles a los sesgos.[40] Asimismo, , respalda la supervisión continua tras la puesta en marcha para detectar riesgos emergentes, tanto técnicos como relacionados con los sesgos, y responder de forma proactiva. Mediante la aplicación de esta política, los hospitales pueden reforzar la resiliencia de los sistemas, salvaguardar la equidad en la toma de decisiones respaldada por la IA y mantener el cumplimiento de la Ley de IA (artículos 9 a 15) y de los requisitos del MDR y el IVDR.

- **Un «expediente del sistema de IA» por herramienta.** Este documento es un expediente estandarizado que se elabora para cada sistema de IA antes de su implantación en una organización sanitaria. El expediente debe resumir la finalidad prevista del sistema, su integración en el flujo de trabajo clínico, los resultados de la validación y los indicadores de rendimiento, prestando especial atención al rendimiento en las distintas poblaciones de pacientes diferenciadas por sexo y género u origen racial o étnico, así como a cualquier disparidad conocida o medida de mitigación. Debe documentar los mecanismos de supervisión humana, aclarar qué decisiones siguen estando exclusivamente bajo el control del personal clínico y señalar cualquier contraindicación relevante para la seguridad y la equidad de los pacientes. Al proporcionar una visión general formalizada y consciente de los sesgos del sistema de IA, el expediente fomenta la transparencia, la gobernanza interna y la supervisión regulatoria, garantizando

[40] Medical Device Coordination Group, *MDCG 2019-16 Rev. 1 - Guidance on Cybersecurity for Medical Devices*, Brussels: European Commission, 2020.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

que las decisiones previas a la implantación tengan en cuenta explícitamente los riesgos de resultados desiguales.

- **Un «paquete combinado de FRIA y DPIA».** Los hospitales públicos que implementan sistemas de IA de alto riesgo se ven cada vez más obligados a evaluar los riesgos para los derechos fundamentales, así como los riesgos para la protección de los datos personales. La integración de estas evaluaciones en una única plantilla que cumpla con las disposiciones de la Ley de IA (art. 27), el RGPD y la DPIA puede reducir la carga administrativa, al tiempo que garantiza que se aborden de forma sistemática cuestiones clave sobre las poblaciones afectadas, los posibles daños, las estrategias de mitigación y los mecanismos de reclamación, prestando especial atención a las posibles desigualdades que afecten a las poblaciones de pacientes diferenciadas por sexo y género y por raza/etnia.

  El «paquete combinado de FRIA y DPIA» funciona principalmente como un instrumento de auditoría ex ante y actualizado periódicamente, situado en la fase previa a la implantación y dentro del ámbito de la auditoría de equidad y derechos. Estas evaluaciones son valoraciones organizadas y documentadas de los riesgos para los derechos fundamentales y la protección de datos, y funcionan como puntos de control de gobernanza más que como herramientas de seguimiento continuo. Si bien deben actualizarse si cambian los riesgos o las condiciones del sistema, su función principal es proporcionar una evaluación sistemática y revisable de los posibles impactos antes y durante la implementación, respaldando el cumplimiento tanto de la Ley de IA como del RGPD, y ayudando a identificar y mitigar los sesgos relacionados con el género o la raza.

  Para reforzar el aspecto de los derechos humanos, debería preverse una consulta significativa con las organizaciones de la sociedad civil (OSC) en el diseño de la evaluación de impacto, a fin de incorporar su experiencia en materia de derechos. Por ejemplo, la Red Europea de Instituciones Nacionales de Derechos Humanos (ENNHRI) insta a la Comisión Europea a complementar las plantillas con orientaciones y consultas públicas para garantizar unas FRIA eficaces que tengan en cuenta la equidad en las dimensiones de género y racial/étnica.[41]

- **Una «política de control de cambios».** Esta política regula las actualizaciones de los sistemas de IA, incluyendo el reentrenamiento de modelos, las

[41] European Network of National Human Rights Institutions (ENNHRI), *Statement on Ensuring Effective Fundamental Rights Impact Assessments (FRIAs) under the EU AI Act*, Brussels: ENNHRI, 2025.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

actualizaciones de software o las modificaciones en los flujos de trabajo clínicos. Las organizaciones deben definir claramente qué constituye un «cambio sustancial» que requiera revalidación o aprobación, garantizando que las actualizaciones no introduzcan riesgos imprevistos para la atención al paciente. En particular, los cambios deben evaluarse en cuanto a sus posibles repercusiones en la equidad de género y racial, incluidos los cambios en el rendimiento desglosado o los resultados diferenciales entre poblaciones diferenciadas por características demográficas. Los hospitales también deben garantizar que los médicos y el personal pertinente estén informados de las modificaciones que podrían afectar a los resultados de la atención de diversas poblaciones de pacientes, lo que permite una supervisión adecuada y la adopción de medidas de mitigación. Al integrar las consideraciones de equidad en la gestión del cambio, los hospitales mantienen tanto la seguridad como la equidad de la toma de decisiones respaldada por la IA en la práctica clínica.

- **Un «sistema de gestión de la calidad» (SGC) para los implementadores que también son proveedores**. Otra dimensión importante del modelo se refiere a la IA desarrollada internamente por hospitales o instituciones de investigación. Las organizaciones sanitarias públicas crean o adaptan cada vez más sus propias herramientas de IA, especialmente en los centros médicos académicos. En tales casos, la institución puede actuar efectivamente como fabricante o proveedor, lo que conlleva responsabilidades reglamentarias adicionales. Las directrices europeas sobre productos sanitarios de desarrollo propio aclaran que las instituciones sanitarias deben aplicar sistemas de gestión de la calidad y mantener documentación técnica que demuestre que las herramientas desarrolladas internamente responden a necesidades específicas de los pacientes que no pueden satisfacerse mediante productos comerciales.[42] El SGQ debe garantizar que se evalúen los resultados de la IA en cuanto a diferencias de rendimiento entre las poblaciones de pacientes diferenciadas por sexo y género y origen racial o étnico, que se documenten las estrategias de mitigación y que los mecanismos de supervisión humana preserven explícitamente la seguridad y la equidad de los pacientes. Incluso cuando se aplican exenciones reglamentarias a los dispositivos de desarrollo propio, las obligaciones derivadas de los derechos fundamentales y de los marcos más amplios de gobernanza de la IA siguen siendo plenamente pertinentes, lo que convierte al sistema de gestión de la calidad en una herramienta fundamental para integrar la equidad en el diseño, la validación y el despliegue de los sistemas de IA biomédica.

[42] Medical Device Coordination Group, *MDCG 2023-1 – Guidance on Health Institution Exemption (MDR/IVDR)*, Brussels: European Commission, 2023.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Fase de seguimiento, auditoría y rendición de cuentas

- **Un «informe de rendimiento y equidad desglosado por población».** Dado que las variaciones en el rendimiento de los algoritmos pueden contribuir a resultados sanitarios desiguales, los hospitales no solo deben exigir a los proveedores que faciliten la información pertinente antes de la adquisición, sino que también deben elaborar análisis internos que muestren el rendimiento de los sistemas en pacientes con características demográficas diversas. Dichos análisis deben centrarse en el rendimiento estratificado por sexo y género y en las diferencias raciales o étnicas pertinentes, siempre que sea lícito y factible, así como en la justificación de cualquier diferencia de rendimiento inevitable. Cuando existan brechas de rendimiento, las organizaciones deben documentar estrategias de mitigación y supervisar la eficacia de dichas medidas a lo largo del tiempo. Este enfoque refleja las expectativas emergentes en la normativa europea sobre IA de que las organizaciones evalúen y mitiguen activamente los riesgos de sesgo en los conjuntos de datos y en los resultados de los modelos. La Ley de IA (art. 10) exige explícitamente el examen y la mitigación de los riesgos de sesgo en los conjuntos de datos, y permite el tratamiento limitado de datos de categorías especiales para la detección de sesgos bajo estrictas garantías.

  El «informe de rendimiento y equidad desglosado por población» funciona principalmente como una herramienta de seguimiento continuo dentro del marco de gobernanza. Requiere una recopilación continua de datos, un seguimiento periódico del rendimiento del sistema en los grupos demográficos pertinentes y una actualización constante de las medidas de mitigación cuando se identifiquen resultados diferenciales. Esto permite a las organizaciones sanitarias detectar y abordar los riesgos de sesgo a lo largo del tiempo. Al mismo tiempo, las conclusiones clave de este informe, como las disparidades identificadas y la eficacia de las estrategias de mitigación, deben sintetizarse e incluirse en la «auditoría de equidad y derechos», garantizando la transparencia y la rendición de cuentas ante la ciudadanía.

- **Un régimen de auditoría de dos niveles que comprenda una auditoría técnica y una auditoría de equidad y derechos.**
  Un régimen de auditoría viable para los hospitales debe equilibrar el rigor con la viabilidad y ser comprensible tanto para el personal como para el público. En este contexto, una auditoría es una revisión formal y periódica del cumplimiento, el rendimiento y los efectos de un sistema de IA sobre los derechos, diseñada para complementar la supervisión continua en lugar de sustituirla.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Auditoría técnica: esta fase consiste en una revisión operativa confidencial y abarca comprobaciones detalladas de carácter interno, entre las que se incluyen pruebas de seguridad, evaluaciones de solidez, validación del rendimiento clínico y cumplimiento de la gobernanza de datos. Debe prestarse especial atención al rendimiento en función de las diferencias de sexo y género, así como raciales o étnicas, evaluando si los resultados del modelo perjudican sistemáticamente a alguna población. Estas actividades deben ajustarse a las disposiciones pertinentes de la Ley de IA (artículos 9 a 15), el MDR y el IVDR, cuando proceda.

Auditoría de equidad y derechos: este segundo nivel se centra en la transparencia, la rendición de cuentas y la mitigación de sesgos. Debe documentar las desigualdades en los resultados, las poblaciones incluidas en las pruebas, las limitaciones conocidas y los mecanismos de reclamación, con un análisis explícito de los impactos de género y raciales. También debe recopilarse evidencia de que la supervisión humana es significativa y capaz de corregir los resultados sesgados. Siguiendo el enfoque de «auditoría y divulgación de sesgos» utilizado en la normativa laboral,[43] un resumen dirigido al público en general puede comunicar las conclusiones clave, las estrategias de mitigación y las mejoras implementadas, sin revelar información privada o confidencial. El «paquete combinado de FRIA y DPIA» también se sitúa dentro del nivel de auditoría de equidad y derechos, reforzando la evaluación rigurosa de los derechos fundamentales, la protección de datos y la equidad demográfica.

- **Un «registro de vigilancia de la IA».** El seguimiento posterior a la implementación también se ve respaldado por la creación de un registro de vigilancia de la IA, inspirado en los sistemas de farmacovigilancia utilizados en la regulación de los medicamentos.[44] Los profesionales sanitarios deberían poder notificar incidentes, cuasi-incidentes o resultados inesperados relacionados con decisiones respaldadas por la IA, con campos explícitos que recojan cualquier posible disparidad observada en materia de sexo y género o de origen racial o étnico. Los informes deben analizarse sistemáticamente para detectar patrones emergentes, incluidos los impactos diferenciales en grupos demográficos específicos, lo que permite a las organizaciones detectar señales de seguridad, problemas de equidad o tendencias de sesgo. Cuando sea necesario, los incidentes, en particular aquellos que indiquen un trato desigual o un riesgo para los grupos protegidos, deben comunicarse sin demora a los proveedores del

[43] New York City Department of Consumer and Worker Protection, *Automated Employment Decision Tools (AEDT)*, New York: City of New York, 2025.

[44] European Medicines Agency, *GVP Module VI – Collection and Submission of Suspected Adverse Reactions (Rev 2)*, London: EMA, 2017.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

sistema, a los comités de gobernanza hospitalaria pertinentes y a las autoridades reguladoras. El registro también debe servir de base para auditorías periódicas de equidad y derechos, lo que favorece la mejora continua y la mitigación del sesgo en el despliegue clínico de la IA.

- **Colaboración a nivel de red e intercambio de incidentes.** La red europea de organizaciones de la sociedad civil (OSC) y hospitales públicos desarrollada en el marco de AEQUITAS funciona como una capa de gobernanza compartida, proporcionando un vocabulario común y normas de notificación para incidentes y hallazgos relacionados con los sesgos de género y raciales en los sistemas de IA. Esta red fomenta el aprendizaje entre centros y la difusión de buenas prácticas para detectar, mitigar y subsanar las desigualdades, de forma similar a los enfoques utilizados en sectores de alta seguridad y en la farmacovigilancia. Al vincular los datos sobre incidentes de sesgo, el rendimiento desglosado y las estrategias de mitigación, la red permite a los hospitales y a las OSC identificar patrones sistémicos de desigualdad y coordinar intervenciones específicas. Dado que el EHDS facilita tanto el uso primario (prestación de asistencia sanitaria) como el uso secundario (investigación y políticas) en un marco regulado, la red puede adaptar sus prácticas de intercambio de datos y aprendizaje a los plazos y actos de aplicación del EHDS a medida que estos maduran, lo que favorece un despliegue más equitativo y transparente de la IA en todas las instituciones.
- **Módulo de formación para profesionales clínicos y expertos de las OSC.** Para mitigar el sesgo de automatización y las desigualdades en la IA clínica, el modelo regulador de AEQUITAS incluye una formación obligatoria diseñada para dotar a los profesionales sanitarios de las habilidades necesarias para comprender las limitaciones de los sistemas de IA, interpretar los resultados de forma crítica y anular o detener los sistemas cuando los resultados puedan producir consecuencias sesgadas o desiguales. La formación hace hincapié en la concienciación sobre los riesgos de sesgo de género y racial, los principios de los derechos fundamentales y las consideraciones éticas, entre ellas la equidad, la no discriminación y la inclusión de los pacientes.

  Este módulo podría considerarse una herramienta facilitadora que apoya activamente el uso del modelo regulador, al garantizar que el personal responsable de la supervisión de la IA pueda aplicar de forma competente los procedimientos de seguimiento, validación y rendición de cuentas desde una perspectiva de equidad. La formación incluirá ejercicios basados en casos prácticos sobre la detección y mitigación de las disparidades de sexo y género, así como raciales y étnicas.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Conclusión

La IA ofrece importantes oportunidades para mejorar los resultados sanitarios, pero sin una gobernanza cuidadosa puede reproducir o agravar las desigualdades de sexo y género y raciales o étnicas, así como otras formas de sesgo sistémico. El modelo regulador propuesto en este documento de políticas traduce los principios de los derechos fundamentales en salvaguardias concretas a lo largo de todo el ciclo de vida de la IA, desde la adquisición y la validación hasta el despliegue, el seguimiento y la rendición de cuentas, prestando especial atención a la detección y mitigación de las diferencias en los resultados entre las distintas poblaciones de pacientes. Mediante la implementación de una arquitectura de gobernanza de la IA estructurada en torno a un conjunto claro de etapas del ciclo de vida, políticas, herramientas prácticas y recursos, los hospitales y las organizaciones de la sociedad civil pueden evaluar, mitigar y documentar de forma sistemática los riesgos, incluidos los impactos desiguales en las poblaciones de pacientes marginadas. Al integrar las lecciones aprendidas en distintos sectores, los requisitos legales de la UE y los marcos de gobernanza internacionales, el modelo regulador ayuda a los profesionales sanitarios a gestionar de forma responsable los sistemas de IA y a garantizar una asistencia sanitaria equitativa, segura y transparente para todos los pacientes, independientemente de su sexo y género, origen racial o étnico u otras características.

Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA se hacen responsables de los mismos. Código del proyecto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI