

05/

Experiencias



05/1

Algunos retos y soluciones para una Inteligencia Artificial más ética y fiable en la sanidad

Carina Dantas.

CEO. SHINE 2Europe. Coimbra (Portugal).

Miriam Cabrita.

Gestor de proyectos. SHINE 2Europe. Coimbra (Portugal).

Luís Midão.

Gestor de proyectos. Centro de Investigação em Tecnologias e Serviços de Saúde (CINTESIS). Porto (Portugal).

Luís Midão.

Profesor. Centro de Investigação em Tecnologias e Serviços de Saúde (CINTESIS). Porto (Portugal).



Las cuestiones éticas relacionadas con la Inteligencia Artificial (I.A.) son un tema emergente y en constante evolución. Ocupan un lugar central en la agenda política europea, pero no ha sido fácil alcanzar un amplio consenso sin largas y difíciles negociaciones.

Los retos éticos a los que hay que hacer frente son complejos y multidisciplinares, por lo que su clasificación facilita un análisis detallado. A partir del trabajo realizado sobre el uso de soluciones digitales basadas en la Inteligencia Artificial en la atención sanitaria, mediante una revisión bibliográfica, talleres y entrevistas con expertos, profesionales y pacientes, se proponen las siguientes categorías: (1) Retos individuales (humanos); (2) Retos técnicos; (3) Retos organizativos; y (4) Retos sociales.

Se necesita un modelo global que analice los retos éticos desde la ideación de cualquier solución tecnológica y los aborde de forma iterativa, con una amplia participación de las distintas partes interesadas, para promover la inclusión, la igualdad de acceso y un uso más justo de la Inteligencia Artificial en la sanidad.

1/

Introducción.

La Inteligencia Artificial (I.A.) trae consigo la expectativa de mejoras en la vida cotidiana de los ciudadanos, prometiendo remodelar profundamente la industria, la sanidad y la sociedad en su conjunto.

Desde el tratamiento médico a la ciberseguridad, pasando por los vehículos autónomos y las decisiones jurídicas, se espera que la I.A. aumente la eficiencia automatizando procesos y mejorando la toma de decisiones, acelerando el análisis y la evaluación de grandes volúmenes

de datos. Sin embargo, a medida que estas tecnologías se desarrollan, prueban y aplican, surgen inquietudes sobre sus implicaciones éticas, sociales y jurídicas.

Las convenciones y normas sociales establecen expectativas predefinidas; éstas, junto con los valores, principios, bagaje cultural, tradiciones y experiencias vitales personales, contribuyen a conformar la visión del mundo de un individuo, interiorizando determinados supuestos o prejuicios. La I.A. no existe en el vacío, sino que se construye a partir de algoritmos creados por humanos, transponiendo así estos prejuicios y reproduciendo el juicio humano. Comprender las fuentes de los sesgos algorítmicos y de los datos permite diversificar los conjuntos de datos; ser más consciente de los posibles sesgos sociales e individuales puede ayudar a mitigarlos a la hora de desarrollar soluciones de I.A.

Aunque algunas de estas desigualdades son sistémicas, por ejemplo, las diferencias socioeconómicas, los prejuicios humanos también desempeñan un papel importante. Una de las causas más comunes de la falta de equidad en la I.A. utilizada en el ámbito médico es el sesgo de los datos utilizados para entrenar los algoritmos (Ricci Lara et al., 2022). Como resultado, muchas de las cohortes clínicas y de investigación existentes que pueden utilizarse para entrenar algoritmos de I.A. médica ya incorporan una representación errónea en los datos.

Los datos de entrenamiento pueden no estar completos, no ser representativos de la población real o incluir otros sesgos relevantes debidos a los parámetros establecidos para la recopilación de datos. Por ejemplo, la mayoría de los conjuntos de datos sobre cáncer se refieren a pacientes mayores de 60 años. Sin embargo, en los últimos años se ha diagnosticado cáncer a un gran número de pacientes más jóvenes, lo que implica que los conjuntos de datos existentes pueden no corresponder a la población real actual (véase el ejemplo del cáncer de intestino en Davis, 2018). Al mismo tiempo, las comunidades LGBTIQ+ no

se tienen en cuenta específicamente en la vigilancia del cáncer, lo que se traduce en una escasez de información sobre la supervivencia al cáncer de estas poblaciones (Boehmer et al., 2011). Este tipo de prejuicios también se observa en la I.A. utilizada para detectar y tratar problemas de salud como el cáncer de piel. Las imágenes de pacientes blancos se utilizan predominantemente para entrenar algoritmos de detección de melanomas, lo que puede dar lugar a peores resultados para las personas de raza negra (Kleinberg et al., 2022).

Sin embargo, los retos éticos no se limitan a los datos. Los sesgos cognitivos (Hardman, 2010) entran en la categoría de prejuicios humanos o sociales y se refieren a los sentimientos hacia una persona o un grupo de personas basados en la percepción de su pertenencia a un grupo. Una lista de más de 180 prejuicios humanos, definidos y clasificados por psicólogos, muestra lo que puede afectar a los individuos cuando tienen que tomar una decisión y, por tanto, puede infiltrarse en los algoritmos de aprendizaje automático cuando se diseñan.

Un estudio de Cowgill et al. (2020) concluyó que

«Cuanto más homogéneo es el equipo [de ingeniería], más probable es que aparezca un error de predicción determinado».

La falta de diversidad, multidisciplinariedad y educación ética en los equipos de desarrollo puede crear una falta de sensibilidad moral y de conciencia hacia las personas que se enfrentan a problemas de discriminación, lo que conduce a una introducción inconsciente de prejuicios en los algoritmos.

Se trata, por tanto, de un área compleja, y esta complejidad aumenta cuando se analizan las diferentes áreas o sectores de actividad en su granularidad, ya que muchos retos éticos son

LH n.339

transversales, pero otros son muy específicos de cada área. Aunque hasta el momento se ha trabajado mucho en diversas recomendaciones de alto nivel, se necesitan herramientas coherentes y armonizadas que puedan aplicarse a los ámbitos específicos de cada solución de Inteligencia Artificial, especialmente a aquellos que implican mayores riesgos de causar daños a los ciudadanos.

Para abordar estos desafíos, la Comisión Europea creó en 2018 el [Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial \(AI HLEG\)](#) con el fin de establecer requisitos generales para una I.A. digna de confianza en Europa. Este Grupo publicó una lista de verificación de evaluación (Comisión Europea. Dirección General de Redes, Contenidos y Tecnologías de la Comunicación - CNECT -, 2020) que proporciona directrices y preguntas para fomentar el desarrollo de la I.A. hacia una mayor fiabilidad. Aunque esta lista de comprobación es un recurso valioso, sus principios son genéricos para todos los ámbitos de la I.A. y la sociedad, y su aplicación en diferentes áreas, en particular la sanidad, requiere marcos y soluciones adaptados.

En este contexto, para garantizar una mayor aceptación y uso de las herramientas de I.A. en el sistema sanitario del futuro, se necesitan nuevos enfoques para aumentar la concienciación y la inclusión, así como métodos para validarlas y aplicarlas en plena consonancia con la diversidad de necesidades y contextos del mundo real.

2/

Resultados.

A partir del trabajo de los autores, es decir, el análisis de la bibliografía existente, combinado con las opiniones de expertos en ética, regula-

ción, ciencias políticas, profesionales sanitarios y tecnológicos, así como pacientes y asociaciones de pacientes, fue posible resumir algunas de las cuestiones más relevantes que deben analizarse en relación con los retos éticos relacionados con el uso de la inteligencia artificial en soluciones sanitarias y las posibles formas de mitigarlos.

Muchos estudios se basan en muestras sesgadas de pacientes y pequeños conjuntos de datos de una sola organización, sin validación externa ([Bhowmik & Eskreis-Winkler, 2022](#)), lo que lleva a cuestionar la precisión de las herramientas de I.A. En este sentido, las bases de datos compartidas entre instituciones pueden ayudar a resolver algunas de estas dificultades, siempre que se garantice la privacidad del paciente ([Houssami et al., 2019](#)).

Además, muchos estudios sobre herramientas basadas en la inteligencia artificial son retrospectivos y carecen de poder predictivo ([Gong et al., 2022](#); [Liang et al., 2022](#); [Vedantham et al., 2023](#)). Se necesitan estudios prospectivos multicéntricos a gran escala con datos exhaustivos y directrices de evaluación de la calidad para avanzar hacia la integración en la práctica clínica ([Bhowmik y Eskreis-Winkler, 2022](#); [Malherbe, 2021](#)).

Sin embargo, la creciente abundancia de datos que se requiere también está planteando retos cada vez mayores a los sistemas sanitarios. A menudo, los centros sanitarios no disponen de un número adecuado de expertos formados para anotar estos datos, lo cual es esencial para los modelos de I.A. (como se explica en [Goisau y Cano Abadía, 2022](#)). Tanto los conjuntos de datos de entrenamiento como los de prueba requieren una anotación meticulosa y el etiquetado de metadatos, utilizando formatos estandarizados para permitir la evaluación y optimización de la equidad algorítmica ([Chua et al., 2021](#)). Por otro lado, garantizar la total privacidad de los datos de los pacientes es una prioridad clara, tanto a efectos de cumplimiento legal como ético. Sin embargo, anonimizar completamente los datos es un reto importante y a menudo sig-

nifica que los resultados obtenidos por la I.A. no pueden compararse con la evolución clínica del paciente, lo que obstaculiza la investigación y las oportunidades de seguir aprendiendo. El aprendizaje federado es uno de los enfoques utilizados actualmente para eludir las restricciones de privacidad (Kandati & Anusha, 2023), ya que permite el entrenamiento colaborativo de algoritmos sin compartir historiales médicos confidenciales, puesto que los datos permanecen almacenados en los servidores originales.

Esta necesidad de grandes conjuntos de datos y de profesionales altamente cualificados preocupa a los sistemas sanitarios, especialmente cuando se trata de herramientas patentadas basadas en la inteligencia artificial, ya que estas solo pueden estar al alcance de instituciones con mayores recursos financieros.

Esto podría agravar las desigualdades en el acceso a la mejor atención sanitaria, como destacan Goisau y Cano Abadía (2022). Además, la falta de análisis coste-efectividad y de investigación basada en la evidencia sobre las herramientas de I.A. hace que no sean fáciles de incluir en los planes de reembolso y en los paquetes de seguros. Según los expertos consultados, para ganarse la confianza de los pagadores, una herramienta basada en la I.A. debe ser más eficiente, tener menos errores y ser capaz de trabajar 24 horas al día, en comparación con la mano de obra humana.

Además, los gastos asociados a las herramientas basadas en I.A. tienen el potencial de aumentar la disparidad entre las personas más ricas y aquellas con menores ingresos, si los conjuntos de datos utilizados para entrenar estas herramientas no incluyen datos de personas con un estatus socioeconómico más bajo, como señalan Cui et al. (2019), y por lo tanto no consideran variables potencialmente relevantes para las personas en situación de vulnerabilidad.

Otro reto emergente es la inclusión de la I.A. en la aplicación nacional del Reglamento sobre los Productos Sanitarios (Radley-Gardner et al., 2016).

Los procedimientos no son uniformes en toda Europa y los equipos de investigación carecen de directrices y procedimientos claros sobre este tema.

También desde el punto de vista de la investigación, la mayoría de los formularios de consentimiento informado obtenidos por las organizaciones sanitarias en relación con los datos de los pacientes no incluyen específicamente la formación en algoritmos, ya que la mayoría de estos consentimientos se recogieron hace mucho tiempo y no constituían una necesidad clara en ese momento. ¿Significa esto que los datos no pueden utilizarse para este fin? ¿Existe alguna excepción de interés público que se aplique a pesar de la falta de consentimiento claro para este fin? (Reddy et al., 2020). En otro aspecto relacionado con el consentimiento, autores como Hickman et al. (2021) plantearon la cuestión de si

“Antes de realizar cálculos como el riesgo de desarrollar una enfermedad, el paciente debería tener que aprobar este análisis tras recibir asesoramiento por parte de un profesional sanitario, de forma similar a los procedimientos previstos actualmente para las pruebas genéticas”.

¿Debería tenerse en cuenta la opinión del paciente a la hora de utilizar este tipo de herramientas? Todas estas cuestiones abiertas, así como los aspectos normativos relacionados, son por tanto ámbitos clave para seguir investigando y elaborando políticas.

El uso de herramientas de I.A. también plantea la cuestión de la responsabilidad (D'Amore et al., 2021; Houssami et al., 2019; Pesapane et al., 2022): ¿quién debe ser considerado responsable en caso de errores de clasificación? ¿Son los algoritmos comprensibles, incluso para profesionales experimentados, como señalan Iqbal y sus colegas (2021)?

Se espera que una herramienta de I.A. proporcione asesoramiento y que la decisión final recaiga en la experiencia del profesional sanitario. Sin embargo, pueden darse situaciones en las que la tecnología sea tan precisa y fiable que la confianza en sus datos resulte inevitable. Esto pone de relieve la compleja relación entre confianza y responsabilidad en el contexto de la asistencia sanitaria basada en la I.A.

Por ello, la madurez de la propia tecnología desempeña un papel importante a la hora de asignar responsabilidades (Retos de la responsabilidad en las tecnologías médicas de IA, 2022). Aunque la legislación vigente intenta identificar a los posibles responsables de predicciones “erróneas” (programadores, médicos, proveedores de hardware), la cuestión sigue siendo ambigua, especialmente con modelos de inteligencia artificial cada vez más complejos (Schneeberger et al., 2020).

Para navegar por este panorama en evolución, son importantes varias estrategias clave. En primer lugar, contar con una legislación nacional e internacional adecuada y/o directrices hospitalarias que se actualicen a un ritmo más rápido y estén vinculadas a resultados basados en la evidencia.

Además, promover enfoques colaborativos de toma de decisiones que incluyan a médicos, pacientes y otros profesionales relevantes puede mitigar los riesgos y garantizar la responsabilidad compartida (Hamilton et al., 2016). Siempre que sea posible, dar a los pacientes la oportunidad de participar activamente en las decisiones de tratamiento fomenta la confianza y la transparencia durante todo el proceso.

Aplicando estas medidas, podemos garantizar que las herramientas de I.A. no sólo mejoren los resultados clínicos, sino que también funcionen dentro de un marco de principios éticos y jurídicos claros.

3/

Discusión.

Es fundamental garantizar la diversidad en todo el proceso de desarrollo de la I.A., desde la concepción hasta la aplicación. Los desarrolladores, los profesionales sanitarios y los usuarios finales deben recibir información, formación y conocimientos adecuados para identificar y mitigar posibles sesgos. Esto subraya la importancia de desarrollar materiales educativos y de formación para todos los grupos de partes interesadas, promoviendo la alfabetización en I.A. y garantizando que los profesionales sanitarios (Lokaj et al., 2023), los pacientes y otras partes interesadas pertinentes tengan los conocimientos y las habilidades necesarios para participar eficazmente y beneficiarse de las soluciones basadas en I.A. implementadas en la práctica diaria (Lekadir et al., 2022).

Los debates en la literatura sobre el sesgo en la I.A. a menudo ignoran sus raíces más profundas y sus efectos más amplios. Cuando el sesgo se considera únicamente un problema técnico con una solución técnica, se pasan por alto aspectos importantes. Por lo tanto, es esencial incluir puntos de vista de las ciencias sociales cuando se exploran las preocupaciones éticas y sociales asociadas al uso de la I.A. (Goisauf & Cano Abadía, 2022; Jongsma & Bredenoord, 2020).

Una cuestión relevante es la necesidad de garantizar que las predicciones proporcionadas por las soluciones basadas en la I.A. sean comprensibles, no sólo para los profesionales sanitarios, sino también para los pacientes y cuidadores. Para generar confianza en los pacientes, es crucial proporcionar información sobre la eficacia y el rendimiento general de la herramienta (por ejemplo, incluyendo información sobre la sensibilidad y la especificidad). Como se destaca en el estudio de Fulmer & Dirks (2018), la confianza abarca varias dimensiones: la confianza del hospital para adquirir la herramienta orientada a la

I.A., la confianza en la tecnología por parte de la industria que la desarrolla, la confianza de los médicos para utilizarla y, por último, pero no por ello menos importante, también la confianza del paciente.

En lo que respecta a los pacientes, la situación es aún más compleja, ya que su confianza en las herramientas de I.A. se ve a menudo influida por su confianza en el hospital e incluso en un médico concreto. Sin embargo, ¿puede la dinámica de poder entre médico y paciente impedir que los pacientes expresen preocupaciones válidas (Ozdemir y Finkelstein, 2018)? Equilibrar la confianza y el conocimiento siguen siendo cuestiones relevantes en la ética sanitaria.

La validación científica de los algoritmos sigue siendo una preocupación clave para los profesionales sanitarios y asistenciales. La falta de procedimientos de validación sólidos, unida a la falta de alineación con los flujos de trabajo clínicos establecidos, incluida la escasa integración con las herramientas existentes (Tsopra et al., 2021), plantea problemas. Además, la ausencia de mecanismos de certificación adecuados complica aún más las cosas cuando se trata de herramientas basadas en la I.A.

Para fomentar la confianza entre los profesionales sanitarios, una estrategia puede ser proporcionar documentación de seguimiento sobre el rendimiento, la usabilidad y la integración de la herramienta basada en la I.A. en las prácticas asistenciales de otras organizaciones sanitarias. Aunque los profesionales sanitarios no necesitan profundizar en detalles técnicos complejos, deben ser capaces de entender cómo los modelos han llegado a predicciones específicas.

Proporcionar un porcentaje de precisión para cada predicción y explicar el razonamiento que la sustenta son pasos cruciales. Al fin y al cabo, los pacientes, sobre todo después de un diagnóstico, se encuentran en un estado vulnerable y es posible que sólo asimilen una parte de la información. La claridad y la comunicación eficaz son fundamentales para generar

confianza y garantizar una toma de decisiones informada.

Basándose en todos los elementos recogidos y debatidos, los autores proponen la siguiente categorización de los retos y prejuicios éticos:

1. Desafíos individuales (humanos), como la falta de educación y formación adecuadas, las creencias y los prejuicios individuales.
2. Desafíos técnicos, por ejemplo, en el diseño defectuoso de algoritmos, conjuntos de datos de entrenamiento sesgados e interpretación incorrecta de los resultados.
3. Desafíos organizativos, por ejemplo, falta de diversidad en los equipos, falta de métodos de auditoría, falta de herramientas disponibles.
4. Desafíos societales, como desigualdades sanitarias, discriminación, problemas éticos y legales, falta de regulación adecuada o de políticas públicas.

Cada una de estas categorías puede ampliarse y adaptarse con retos de mayor nivel o más específicos, ajustándola a los casos prácticos concretos que se analicen. Abordar estos retos de forma colectiva es vital para el uso ético y eficaz de la I.A. en la asistencia sanitaria.

Las posibles medidas de mitigación también pueden agruparse en estas mismas cuatro áreas:

Por ejemplo, en lo que respecta a los factores individuales (humanos), es muy importante el desarrollo de materiales de formación y/o educación específicos para cada tipo de destinatario.

En cuanto a los factores técnicos, vinculados principalmente a la industria, debe hacerse énfasis en la mejora de la solidez de los resultados de la Inteligencia Artificial, que necesitan conjuntos de datos fiables, de buena calidad y de gran tamaño. La auditoría y supervisión periódicas de los algoritmos es esencial para garantizar un buen rendimiento a medio y largo plazo.

LH n.339

En cuanto a los aspectos organizativos, asegurar equipos multidisciplinares con la formación adecuada y aplicar métodos que garanticen la participación de los ciudadanos y las organizaciones de la sociedad civil en el diseño y desarrollo de soluciones de I.A. es clave para superar los prejuicios y retos detectados.

Las medidas societales abarcan desde la participación de las partes interesadas hasta los aspectos políticos, éticos, jurídicos y normativos. Los métodos procedentes de la innovación social y las ciencias sociales, combinados con los conocimientos técnicos y clínicos, son útiles y deben seguir investigándose y aplicándose con vistas a una sociedad más integradora.

4/

Conclusión.

Del análisis realizado se ha podido concluir que los retos a los que se enfrenta el desarrollo de herramientas de I.A. en el ámbito sanitario son multifacéticos y deben ser estructurados y analizados cuidadosamente.

El modelo propuesto en este artículo para agrupar los retos y sesgos éticos del uso de la I.A. en la sanidad pretende facilitar el análisis y la resolución de estos retos hacia una sociedad más inclusiva, justa y digna de confianza.

Bibliografía

▶ **Bhowmik, A., & Eskreis-Winkler, S. (2022).** *Deep learning in breast imaging.* BJR Open, 4(1), 20210060. <https://doi.org/10.1259/bjro.20210060>

▶ **Boehmer, U., Miao, X., & Ozonoff, A. (2011).** *Cancer survivorship and sexual orientation.* Cancer, 117(16), 3796–3804. <https://doi.org/10.1002/cncr.25950>

▶ **Chua, I., Gaziel-Yablowitz, M., Korach, Z., Kehl, K., Levitan, N., Arriaga, Y., Jackson, G., Bates, D., & Hassett, M. (2021).** *Artificial intelligence in oncology: Path to implementation.* Cancer Medicine, 10(12), 4138–4149. <https://doi.org/10.1002/cam4.3935>

▶ *Coordinated Plan on Artificial Intelligence 2021 Review (Shaping Europe's Digital Future).* (2021). **European Commission.** <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

▶ **Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020).** *Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics* (arXiv:2012.02394). arXiv. <https://doi.org/10.48550/arXiv.2012.02394>

▶ **Cui, C., Chou, S.-H. S., Brattain, L., Lehman, C. D., & Samir, A. E. (2019).** *Data Engineering for Machine Learning in Women's Imaging and Beyond.* American Journal of Roentgenology, 213(1), 216–226. <https://doi.org/10.2214/AJR.18.20464>

Davis, N. (2018, October 22).

Bowel cancer on the rise among young people in Europe.

The Guardian.

<https://www.theguardian.com/society/2018/oct/22/bowel-cancer-on-the-rise-among-young-people-in-europe>

European Commission. Directorate General for Communications Networks, Content and Technology. (2020).

The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment.

Publications Office.

<https://data.europa.eu/doi/10.2759/791819>

Fulmer, A., & Dirks, K. (2018).

Multilevel trust: A theoretical and practical imperative.

Journal of Trust Research, 8(2), 137–141.

<https://doi.org/10.1080/21515581.2018.1531657>

Goisau, M., & Cano Abadía, M. (2022).

Ethics of AI in Radiology:

A Review of Ethical and Societal Implications.

Frontiers in Big Data, 5, 850383.

<https://doi.org/10.3389/fdata.2022.850383>

Gong, X., Guo, Y., Zhu, T., Peng, X., Xing, D., & Zhang, M. (2022).

Diagnostic performance of radiomics in predicting axillary lymph node metastasis in breast cancer: A systematic review and meta-analysis.

Frontiers in Oncology, 12. Scopus.

<https://doi.org/10.3389/fonc.2022.1046005>

Hamilton, D. W., Heaven, B., Thomson, R. G., Wilson, J. A., & Exley, C. (2016).

Multidisciplinary team decision-making in cancer and the absent patient: A qualitative study.

BMJ Open, 6(7), e012559.

<https://doi.org/10.1136/bmjopen-2016-012559>

Hardman, D. (2010).

Judgment and decision making:

Psychological perspectives (Reprinted).

BPS Blackwell.

Hickman, S. E., Baxter, G.

C., & Gilbert, F. J. (2021).

Adoption of artificial intelligence in breast imaging: Evaluation, ethical constraints and limitations.

British Journal of Cancer, 125(1), 15–22.

<https://doi.org/10.1038/s41416-021-01333-w>

Houssami, N., Kirkpatrick-Jones, G.,

Noguchi, N., & Lee, C. I. (2019).

Artificial Intelligence (AI) for the early detection of breast cancer: A scoping review to assess AI's potential in breast screening practice.

Expert Review of Medical Devices, 16(5),

351–362. <https://doi.org/10.1080/17434440.2019.1610387>

Iqbal, M. J., Javed, Z., Sadia, H., Qureshi,

I. A., Irshad, A., Ahmed, R., Malik, K.,

Raza, S., Abbas, A., Pezzani,

R., & Sharifi-Rad, J. (2021).

Clinical applications of artificial intelligence and machine learning in cancer diagnosis: Looking into the future.

Looking into the future.

Cancer Cell International, 21(1), 270.

<https://doi.org/10.1186/s12935-021-01981-1>

Jongsma, K. R., & Bredenoord, A. L. (2020).

Ethics parallel research: An approach for (early) ethical guidance of biomedical innovation.

BMC Medical Ethics, 21(1), 81. <https://doi.org/10.1186/s12910-020-00524-z>

- ▶ **Kandati, D. R., & Anusha, S. (2023).** *Security and privacy in federated learning: A survey.* Trends in Computer Science and Information Technology, 8(2), 029–037. <https://doi.org/10.17352/tcsit.000066>
- ▶ **Kleinberg, G., Diaz, M. J., Batchu, S., & Lucke-Wold, B. (2022).** *Racial underrepresentation in dermatological datasets leads to biased machine learning models and inequitable healthcare.* Journal of Biomed Research, 3(1), 42–47.
- ▶ **Lekadir, K., Giancula, Q., Anna, T. G., & Gallin, C. (2022).** *Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts.* European Parliament. Directorate General for Parliamentary Research Services. <https://data.europa.eu/doi/10.2861/568473>
- ▶ **Liang, X., Yu, X., & Gao, T. (2022).** *Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis.* European Journal of Radiology, 150, 110247. <https://doi.org/10.1016/j.ejrad.2022.110247>
- ▶ **Lokaj, B., Pugliese, M.-T., Kinkel, K., Lovis, C., & Schmid, J. (2023).** *Barriers and facilitators of artificial intelligence conception and implementation for breast imaging diagnosis in clinical practice: A scoping review.* European Radiology. <https://doi.org/10.1007/s00330-023-10181-6>
- ▶ **Malherbe, K. (2021).** *Tumor Microenvironment and the Role of Artificial Intelligence in Breast Cancer Detection and Prognosis.* The American Journal of Pathology, 191(8), 1364–1373. <https://doi.org/10.1016/j.ajpath.2021.01.014>
- ▶ **Ozdemir, S., & Finkelstein, E. (2018).** *Cognitive Bias: The Downside of Shared Decision Making.* JCO Clinical Cancer Informatics, 2. <https://doi.org/10.1200/CCI.18.00010>
- ▶ **Pesapane, F., Trentin, C., Montesano, M., Ferrari, F., Nicosia, L., Rotili, A., Penco, S., Farina, M., Marinucci, I., Abbate, F., Meneghetti, L., Bozzini, A., Latronico, A., Liguori, A., Carrafiello, G., & Cassano, E. (2022).** *Mammography in 2022, from Computer-Aided Detection to Artificial Intelligence Applications.* Clinical and Experimental Obstetrics & Gynecology, 49(11), 237. <https://doi.org/10.31083/j.ceog4911237>
- ▶ **Radley-Gardner, O., Beale, H., & Zimmermann, R. (Eds.). (2016).** *Fundamental Texts On European Private Law.* Hart Publishing. <https://doi.org/10.5040/9781782258674>
- ▶ **Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020).** *A governance model for the application of AI in health care.* Journal of the American Medical Informatics Association, 27(3), 491–497. <https://doi.org/10.1093/jamia/ocz192>

Ricci Lara, M. A., Echeveste, R., & Ferrante, E. (2022).

Addressing fairness in artificial intelligence for medical imaging.

Nature Communications, 13(1), 4581.

<https://doi.org/10.1038/s41467-022-32186-3>

Schneberger, D., Stöger, K., & Holzinger, A. (2020).

The European Legal Framework for Medical AI.

In A. Holzinger, P. Kieseberg, A. M. Tjoa, & E. Weippl (Eds.), *Machine Learning and Knowledge Extraction* (Vol. 12279, pp. 209–226). Springer International Publishing. https://doi.org/10.1007/978-3-030-57321-8_12

Tsopra, R., Fernandez, X., Luchinat, C., Alberghina, L., Lehrach, H., Vanoni, M., Dreher, F., Sezerman, O. U., Cuggia, M., De Tairac, M., Miklasevics, E., Itu, L. M., Geanta, M., Ogilvie, L., Godey, F., Boldisor, C. N., Campillo-Gimenez, B., Cioroboiu, C., Ciusdel, C. F., ... Burgun, A. (2021).

A framework for validating AI in precision medicine: Considerations from the European ITFoC consortium.

BMC Medical Informatics and Decision Making, 21(1), 274. <https://doi.org/10.1186/s12911-021-01634-3>

Vedantham, S., Shazeeb, M. S., Chiang, A., & Vijayaraghavan, G. R. (2023).

Artificial Intelligence in Breast X-Ray Imaging.

Seminars in Ultrasound, CT, and MR, 44(1), 2–7. <https://doi.org/10.1053/j.sult.2022.12.002>

