

ÉTICA DE LA INTELIGENCIA ARTIFICIAL

RAFAEL AMO USANOS
SARA LUMBRERAS SANCHO
ÍÑIGO NAVARRO MENDIZÁBAL
Directores



Dykinson, S.L.

Colección
IA, Robots, y Bioderecho

**ÉTICA DE LA
INTELIGENCIA ARTIFICIAL**

Colección
IA, ROBOTS, Y BIODERECHO

Directores

FRANCISCO LLEDÓ YAGÜE
Catedrático de Derecho Civil de la Universidad de Deusto

IGNACIO BENÍTEZ ORTÚZAR
Catedrático de Derecho Penal de la Universidad de Jaén

CRISTINA GIL MEMBRADO
Catedrática de Derecho Civil de la Universidad de las Islas Baleares

ÓSCAR MONJE BALMASEDA
Catedrático de Derecho Civil de la Universidad de Deusto

Coordinadores

M^a JOSÉ CRUZ BLANCA
Profesor titular de Derecho Penal de la Universidad de Jaén

IGNACIO LLEDÓ BENITO
Profesor Derecho Penal de la Universidad de Sevilla. Profesor titular acreditado (ANECA)

Comité científico

LORENZO MORILLAS CUEVA
Catedrático de Derecho Penal de la Universidad de Granada

MANUEL MARCHENA GARCÍA
Presidente de la Sala Segunda del Tribunal Supremo

PILAR FERRER VANRELL
Catedrática de Derecho Civil de la Universidad de las Islas Baleares

JOSÉ ÁNGEL MARTÍNEZ SANCHIZ
Notario. Académico de "número" de la Real Academia de Legislación y Jurisprudencia

VICTORIO MAGARIÑOS BLANCO
Notario y miembro de la Comisión General de Codificación

PIERRE LLUIGI D'ELLOSSO
Fiscal General de la República Emérito. Fiscal Nacional Antimafia (Italia)

ALICIA SÁNCHEZ SÁNCHEZ
Magistrado-Juez Registro Civil de Bilbao

LUCÍA RUGGERI
Professore ordinario di Diritto privato presso Università degli Studi di Camerino

CARMEN OCHOA MARIETA
Directora médico ura, Cer.Santander,S.L, Medicina de la reproducción

MARIAN M. DE PANCORBO
Catedrática de Biología Celular, Coordinadora Centro de Investigación Lascaray Ikergunea / Lascaray Research Center, Investigadora Principal Grupo biomics / biomics Research Group

LUIS MARTÍNEZ LÓPEZ
Catedrático de Lenguajes y Sistemas Informáticos de la Universidad de Jaén

HUMBERTO NICANOR BUSTINCE SOLA
Catedrático de Ciencias de la Computación e Inteligencia Artificial de la Universidad Pública de Navarra

ÉTICA DE LA INTELIGENCIA ARTIFICIAL

RAFAEL AMO USANOS
SARA LUMBRERAS SANCHO
ÍÑIGO NAVARRO MENDIZÁBAL
(Directores)

Dykinson, S.L.

No está permitida la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea este electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (art. 270 y siguientes del Código Penal).

Diríjase a Cedro (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra. Puede contactar con Cedro a través de la web www.conlicencia.com o por teléfono en el 917021970/932720407

Este libro ha sido sometido a evaluación por parte de nuestro Consejo Editorial

Para mayor información, véase www.dykinson.com/quienes_somos

© Copyright by
Los autores
Madrid

Editorial DYKINSON, S.L. Meléndez Valdés, 61 - 28015 Madrid
Teléfono (+34) 91 544 28 46 - (+34) 91 544 28 69
e-mail: info@dykinson.com
<http://www.dykinson.es>
<http://www.dykinson.com>

ISBN: 979-13-7047-159-0
Depósito Legal: M-9704-2026
DOI: <https://doi.org/10.14679/4908>

ISBN electrónico: 979-13-7047-255-9

Preimpresión por:
Besing Servicios Gráficos S.L.
e-mail: besingsg@gmail.com

Índice

PRÓLOGO	21
JUAN MANUEL GARCÍA	
PRÓLOGO	23
JAIME TATAY, SJ, PhD	
INTRODUCCIÓN GENERAL	25
RAFAEL AMO USANOS - SARA LUMBRERAS SANCHO - IÑIGO NAVARRO MENDIZABAL	
CAPÍTULO 1. FUNDAMENTOS TECNOLÓGICOS DE LA IA	37
ÁLVARO LÓPEZ - SARA LUMBRERAS	
1. Introducción: Un momento sin precedentes	37
2. Cómo funciona la IA.....	39
2.1. Orígenes de la IA y principales corrientes.....	39
2.2. Tipos de aprendizaje automático	41
2.3. Los avances que han hecho posible la situación actual.....	43
2.4. Instruct GPT: un proyecto en el que encajaron todas las piezas	46
3. Capacidades Actuales de la IA.....	48
3.1. Sistema 1: procesamiento rápido, heurístico e implícito....	49
3.2. Sistema 2: razonamiento deliberativo, simbólico o secuencial	51
4. Aplicaciones de la IA	53
5. Los Retos de la IA	55
6. Los mitos de la IA	56
6.1. Mito 1: <i>La IA siempre funciona bien</i>	56
6.2. Mito 2: <i>La IA es una caja negra impenetrable</i>	57
6.3. Mito 3: <i>Si habla como humano, entonces piensa como humano</i>	57
7. La ciencia en la era de la IA.....	58

7.1.	Del experimento controlado al dato masivo: una evolución histórica	59
7.2.	Fundamentos conceptuales de la inferencia causal contemporánea	59
7.3.	El papel de la inferencia contrafactual	60
7.4.	Nuevas herramientas, nuevos riesgos	60
8.	Delegación de decisiones	61
9.	Manipulación	61
10.	Tecnología en las relaciones humanas	62
11.	Privacidad, vigilancia y seguridad	63
12.	Referencias bibliográficas.....	64
CAPÍTULO 2. FUNDAMENTOS FILOSÓFICOS DE LA IA		69
RAFAEL AMO USANOS - SARA LUMBRERAS SANCHO		
1.	Introducción	69
2.	Filosofía de la mente.....	69
2.1.	Taxonomías de la IA	70
2.2.	La IA simbólica y subsimbólica	71
2.3.	Balance filosófico de la IA	73
3.	Epistemología y teoría de la ciencia	75
4.	Filosofía de la tecnología y la IA.....	77
5.	La ética de la IA.....	78
5.1.	Estatuto epistemológico de la ética de la IA	79
5.2.	Panorama de las problemáticas éticas de la IA	82
6.	IA y nuestra comprensión de la naturaleza humana.....	89
7.	Referencias Bibliográficas	90
CAPÍTULO 3. FUNDAMENTOS LEGALES EN LA ÉTICA DE LA IA		95
IÑIGO NAVARRO MENDIZABAL		
1.	Introducción: la intersección entre el Derecho y la Ética en la IA	95
1.1.	La necesaria relación entre normas jurídicas y principios éticos	95
1.2.	La influencia de la ética en la normativa en la UE	98
2.	Comparativa internacional en el enfoque ético y la regulación de la IA...	99

2. 1.	EE.UU: ética del mercado, innovación y regulación fragmentada	100
2.2.	China: soberanía tecnológica, armonía social y control centralizado	105
2. 3.	UE: ética de los derechos fundamentales y gobernanza basada en el riesgo	111
2. 4.	Conclusión general de la comparativa	116
3.	SOFT LAW Y ESTÁNDARES INTERNACIONALES: GOBERNANZA ÉTICA SIN COERCIÓN JURÍDICA	117
3. 1.	Introducción: el rol del <i>Soft Law</i> en la regulación de la IA	117
3. 2.	Marcos éticos multilaterales: UNESCO y OCDE	118
3. 3.	Iniciativas privadas y sectoriales en el ecosistema del <i>Soft Law</i> para la IA	119
3. 4.	Conclusión: El <i>Soft Law</i> como catalizador normativo en la Era de la IA	121
4.	Referencias Bibliográficas	122

CAPÍTULO 4. PRIVACIDAD Y PROTECCIÓN DE DATOS PERSONALES EN EL CONTEXTO DE LA IA

FEDERICO DE MONTALVO JÄÄSKELÄINEN

1.	Big Data e IA: una relación inescindible y circular	129
2.	El derecho a la protección de datos personales	135
3.	Garantías legales de protección de la privacidad y la confidencialidad de los datos	139
4.	¿Es adecuado este marco de garantías asentado, sustancialmente, en el consentimiento informado en el contexto de la IA?	145
5.	La garantía de la seudonimización como modelo superador del binomio consentimiento-anonimización	149
6.	Otras propuestas para la protección de los datos en el contexto de la IA	157
7.	La alfabetización digital como gran reto	159
8.	Referencias bibliográficas	163

CAPÍTULO 5. INTRODUCCIÓN A LA ALGORÉTFICA: TRANSPARENCIA Y RESPONSABILIDAD	167
SARA LUMBRERAS - FRANCESC TORRALBA	
1. Introducción: algoritmos y ética de la IA.....	167
2. El peso de los algoritmos hoy: transparencia y rendimiento.....	168
3. Sesgos inherentes: el problema del sesgo	171
4. Seguridad de los algoritmos.....	172
5. Vulnerabilidad humana ante los algoritmos	173
6. Conclusiones	174
7. Referencias bibliográficas.....	175
CAPÍTULO 6. RESPETO DE LA AUTONOMÍA HUMANA	177
IÑIGO A. NAVARRO MENDIZABAL	
1. Introducción	177
2. Conceptualización de la autonomía personal en el ámbito de la IA	178
3. Dimensiones de la autonomía humana en la IA	181
3.1. Autonomía cognitiva y emocional	182
3.2. Autonomía decisonal y moral	185
3.3. Autonomía informativa	188
4. Repaso de los beneficios y amenazas de la IA para la autonomía hu- mana	190
4.1. Beneficios de la IA para la autonomía humana	191
4.2. Amenazas de la IA para la autonomía humana	194
5. Criterios éticos para una IA respetuosa de la autonomía humana	198
5.1. Criterios para proteger la autonomía cognitiva y emocio- nal	198
5.2. Criterios para preservar la autonomía decisonal y moral .	200
5.3. Criterios para salvaguardar la autonomía informativa	203
6. Referencias bibliográficas.....	204

CAPÍTULO 7. CINCO GRADOS DE AUTONOMÍA ARTIFICIAL EN LA ÉTICA DE LA TOMA DE DECISIONES: DE LA FUNCIÓN AL DESPERTAR DE DILECCIÓN	209
LUIS E. ECHARTE ALONSO	
1. Supervisión plena	209
2. Autonomía principal e instrumental	211
3. Nuevos agentes racionales	213
4. Culminación del proceso de tecnificación	215
5. Mecanización interna y externa de la mente.....	217
6. Mentes colmena.....	219
7. El nuevo ocio y la industria de la autenticidad.....	221
8. Gradualidad de reverso.....	223
9. La holgura del ser.....	224
10. Pesimismo computacional.....	227
11. Estética trascendental	229
12. La tensión creativa de la decisión moral.....	232
13. Referencias bibliográficas.....	236
CAPÍTULO 8. IA Y RELACIONES HUMANAS	239
ANTONIO JESÚS MARÍA SÁNCHEZ ORANTOS - JUAN JESÚS GUTIERRO CARRASCO	
1. Introducción: exigencia de fidelidad a la pretensión de la filosofía primera.....	239
2. El cuerpo humano y la imposible analogía de hardware/software para explicar la vida humana.....	241
3. La vida humana desde la vivencia de la alteridad radical.....	246
4. Conclusión	260
5. Referencias Bibliográfica.....	261
CAPÍTULO 9. DISCERNIR LO JUSTO EN EL ÁMBITO DE LA IA	265
JULIO L. MARTÍNEZ, SJ	
1. Introducción	265
2. Noción básica de justicia y principales hitos en su desarrollo.....	266
3. ¿Qué implica meter la justicia en la consideración de la Ia?	268

4.	Recorrido propuesto	270
5.	La tecnociencia y el utilitarismo	270
6.	La justicia pide libertad con igualdad	273
7.	El enfoque de la justicia centrado en el desarrollo de capacidades	274
8.	Epistemología contextualista: diversos bienes sociales piden distintas esferas de justicia	275
9.	Perspectivas global e inclusiva de la justicia	276
10.	Justicia como participación de todos en la vida de la comunidad: no al descarte	277
11.	El deterioro de los derechos y libertades en la sociedad digital.....	280
12.	«Brechas tecnológicas».....	281
13.	Sostenibilidad e IA	283
14.	Concentración de poder y paradigma tecnocrático	284
15.	Deontologismo más teleologismo con mucho diálogo	285
16.	Referencias Bibliográficas	287
CAPÍTULO 10. ÉTICA DEL USUARIO DE LA IA		291
FRANCISCO JAVIER REAL ÁLVAREZ		
1.	¿Por qué es necesaria una ética del usuario de la IA?.....	293
1.1.	El mundo tecnolíquido como nuevo ecosistema vital	293
1.2.	El tipo de hombre que emerge de este contexto.....	294
1.3.	Consecuencias de la ausencia de una ética del usuario	295
2.	Fundamentos para una ética del usuario	297
2.1.	Carencias de algunas posturas éticas	297
2.2.	El humanismo cristiano como base antropológica	299
2.3.	La ética de las virtudes como camino de humanización y relación	300
3.	Hacia una ética del usuario basada en las virtudes	302
3.1.	El perfil moral del usuario virtuoso.....	302
3.2.	Virtudes clave para el uso ético de la IA	303
3.3.	Formación de la conciencia: la sabiduría digital	306
4.	Conclusión	308
5.	Referencias Bibliográficas	309

CAPÍTULO 11. ÉTICA Y USOS MILITARES DE LA INTELIGENCIA ARTIFICIAL	315
JUAN A. MOLINER GONZÁLEZ	
1. Algunas consideraciones previas sobre guerra y tecnología	315
2. La doctrina de la Guerra Justa y su relación con la IA	317
3. Las aplicaciones militares de la IA	319
4. Principios éticos que deben regir el empleo de la IA en guerras y conflictos armados	321
4.1. La importancia de la discriminación como principio ético y legal	321
4.2. El principio de prevención	322
4.3. El principio ético de la reducción del riesgo a los combatientes propios	323
4.4. Otros desafíos éticos en el uso militar de la IA	324
5. La asunción de responsabilidades, ¿de hombres o máquinas?	325
5.1. La autonomía de los sistemas de armas	325
5.2. Control humano significativo	326
5.3. Predictibilidad, explicabilidad y rendición de cuentas	329
6. ¿El avance de la IA hacia los robots «éticos»?	330
8. Conclusiones	332
9. Referencias bibliográficas.....	335
CAPÍTULO 12. BIOÉTICA, SALUD E INTELIGENCIA ARTIFICIAL	337
FEDERICO DE MONTALVO JÄÄSKELÄINEN - RAFAEL AMO USANOS	
1. La aplicación de la IA a la relación sanitaria como mejora de la asistencia.....	337
2. Aplicaciones de IA en salud.....	339
2.1. IA en diagnóstico	339
2.2. IA en tratamientos	340
2.3. Medicina preventiva y autocuidado	341
2.4. La IA en la investigación biomédica y el descubrimiento de medicamentos	343
2.5. Gestión optimizada de los hospitales y mejora de la calidad asistencial	344
3. Los retos éticos de la IA en el ámbito de la salud.....	345

4.	La corporeidad como límite a la implementación de la IA: más allá de la mera supervisión humana	347
5.	Justicia, salud e IA	359
6.	Sesgos en salud	362
7.	Sostenibilidad e IA	364
	7.1. Aplicaciones en salud de la IA y su relación al ODS 3	364
	7.2. Relación con la sostenibilidad ambiental, social y democrática.....	365
8.	Referencias bibliográficas	366
 CAPÍTULO 13. INTELIGENCIA ARTIFICIAL Y DERECHO EN LA EVOLUCIÓN HACIA UN TIEMPO DE SINGULARIDAD TECNOLÓGICA: FUNDAMENTOS ÉTICOS PARA UNA FUNCIÓN LEGAL EN TRANSFORMACIÓN.....		373
OBSERVATORIO LEGALTECH ICADE-GARRIGUES*		
1.	Introducción	373
2.	Marcos para el análisis de la función legal en la era de la IA avanzada	375
	2.1. Primer marco: clasificación de los servicios legales en función al uso de la tecnología: (1:1, 1:n, 0:n)	375
	2.2. Segundo marco: evolución de la IA (pre-IAGen, IAGen, IAG y SIA)	376
3.	Impactos combinados (contexto tecnológico x época IA) en cada dimensión de la función legal.....	382
	3.1. Administración de Justicia: Principio básico, IA de alto riesgo bajo control humano.....	385
	3.2. Abogacía: de la práctica artesanal a la prestación <i>industrial</i> escalada por IA.....	390
	3.3. Notarios y Registradores: digitalización de la fe pública y desafíos de la automatización.....	394
4.	Principios éticos adaptativos para la función legal en la era de la IA ...	397
	4.1. Dignidad humana y derechos fundamentales.....	398
	4.2. Transparencia y explicabilidad	399
	4.3. Responsabilidad y control humano.....	401
	4.4. Equidad y no discriminación.....	403
	4.5. Vacíos normativos y retos abiertos.....	405
5.	Conclusiones	407
6.	Referencias bibliográficas.....	409

**CAPÍTULO 14. DIFUSIÓN DE HERRAMIENTAS DE IA EN LA EMPRESA:
IMPACTO SOCIAL Y ÉTICO..... 411**

JUAN JUNG - GONZALO GÓMEZ-BENGOECHEA

1.	Motivación y literatura	411
2.	Análisis empírico	413
	2.1. Análisis descriptivo	413
	2.2. El modelo	415
3.	Interpretación de los resultados.....	418
4.	Discusión	420
	4.1. Dilemas éticos	420
	4.2. Políticas públicas para una IA inclusiva	423
	4.3. El futuro del (des)empleo	425
5.	Conclusión	427
6.	Referencias Bibliográficas	428

**CAPÍTULO 15. ÉTICA DE LA IA EN LA GESTIÓN DE RECURSOS
HUMANOS..... 433**

JOSEBA ARANO ECHEBARRIA

1.	Cómo abordar una gestión ética de la IA en las organizaciones desde la función de Personas	433
2.	El impacto de la IA en las personas desde una perspectiva de la gestión de RH	436
3.	LAS PROMESAS DE LA IA PARA LA FUNCIÓN DE PERSONAS	438
4.	Riesgos éticos en la adopción de la IA en la gestión de Personas en las Organizaciones	443
5.	Respuestas desde la función de personas para una gestión ética y excelente de la IA en la organización.....	445
6.	Implicaciones de la adopción de la IA en la función de RH.....	450
7.	Referencias bibliográficas.....	451

CAPÍTULO 16. DEMOCRACIA ARTIFICIAL ¿UN NUEVO MODELO DE DEMOCRACIA?	453
RAFAEL RUBIO NÚÑEZ	
1. Inteligencia Artificial y Democracia: Entre <i>Prometeo, Hermes y Pandora</i>	453
2. Cambios estructurales que provoca la IA en la sociedad. Estado: población, territorio y soberanía	456
3. Riesgos y amenazas de la IA para la democracia.....	457
3.1. Modelos de democracia en tensión: tecno-utopía vs. tecnocracia	458
3.2. Desigualdad tecnológica y brecha de poder	461
3.3. Erosión del Estado de Derecho y derechos fundamentales	463
3.4. Autonomía y manipulación algorítmica	466
3.5. Representación y opinión pública	468
4. Conclusiones: La IA al servicio de la democracia.....	472
5. Referencias bibliográficas.....	474
CAPÍTULO 17. SABIDURÍA PARA GOBERNAR LA INTELIGENCIA: IRRUPCIÓN DE LA IA EN EL SISTEMA INTERNACIONAL E IMPLICACIONES ÉTICAS	479
JAVIER MERCHÁN - JULIA LOGA	
1. Los dilemas de la Inteligencia Artificial	479
2. Pasado y futuro de la Inteligencia Artificial	481
3. Marco conceptual del trinomio: Relaciones Internacionales, Ética e Inteligencia Artificial	483
3.1. De la IA en RRII a la IA como problema ético-político de RRII	483
3.2. Un triángulo analítico: poder, normas y técnica	484
3.3. ¿Qué lectura del trinomio RRII-ética-IA puede hacerse desde las principales teorías de las Relaciones Internacionales?	486
4. Regular bajo un criterio filosófico preciso: la diplomacia de la IA.....	490
5. Observaciones finales: ética y regulación de la IA en perspectiva comparada.....	494
6. Referencias bibliográficas.....	496

CAPÍTULO 18. EDUCACIÓN Y USOS DE LA IA.....	499
FRANCISCO RAMÍREZ FUEYO	
1. Principio de beneficencia	500
1.1. Crecimiento en competencia intelectual y técnica	501
1.2. Creatividad e innovación.....	503
1.3. Amor por el conocimiento y curiosidad intelectual	504
1.4. Responsabilidad profesional	506
1.5. Bienestar físico, emocional, psicológico y espiritual	507
1.6. Desarrollo del sentido estético.....	508
1.7. Capacidad de tener de relaciones profundas y significati- vas, y de comunicación	511
1.8. Responsabilidad pedagógica y evaluación	512
2. Principio de autonomía	514
2.1. Capacidad para decidir y para elegir el propio camino de vida.....	514
2.2. Autonomía personal en el aprendizaje y sistemas autóno- mos de docencia	515
2.3. Pensamiento crítico, formación del juicio moral, hones- tidad académica, confianza, transparencia, asunción de responsabilidades.....	516
2.4. Habilidades para colaborar	518
3. Principio de justicia.....	519
3.1. Solidaridad, sostenibilidad ecológica, participación social....	519
3.2. Equidad, inclusión, cuidado, tolerancia activa y diversidad....	521
4. Referencias bibliográficas.....	523
CAPÍTULO 19. ARTE, LITERATURA Y ENTRETENIMIENTO.....	527
MARÍA LUISA ROMANA GARCÍA	
1. Cultura, sociedad y necesidades humanas.....	527
1.1. La cultura como fenotipo ampliado.....	527
1.2. Interpretación social de las necesidades humanas.....	528
1.3. La creatividad como dimensión humana.....	529
1.4. Funciones sociales del entretenimiento	530
2. Tecnología, cultura y responsabilidad.....	532
2.1. La tecnología como configuradora de la cultura.....	532
2.2. La arrogancia técnica y sus riesgos	532

2.3.	El dilema del dinero	533
3.	Economía, entretenimiento y manipulación.....	534
3.1.	La hegemonía simbólica	534
3.2.	El entretenimiento como herramienta de control: estrés, presión y adicción	534
3.3.	La medición como instrumento de explotación	535
4.	La defensa de la cultura	536
4.1.	Ética en el ecosistema cultural	536
4.2.	Mecanismos preventivos y sancionadores	537
4.3.	Evaluación continua y ciudadanía crítica	538
5.	Conclusiones	538
6.	Referencias bibliográficas.....	540
 CAPÍTULO 20. INVESTIGACIÓN EN SALUD E INTELIGENCIA ARTIFICIAL		543
JULIO C. DE LA TORRE-MONTERO - BLANCA EGGA-ZEROLO		
1.	Introducción	543
2.	Estrategias en Salud Digital y aplicaciones de IA.....	545
3.	Terapias Digitales.....	548
4.	Relación clínica e inteligencias múltiples, inteligencia emocional e IA.....	549
5.	El Cuidado como valor para la Salud	552
6.	Mirada al presente y futuro.....	556
7.	Referencias bibliográficas.....	556
 CAPÍTULO 21. SALUD MENTAL E INTELIGENCIA ARTIFICIAL		561
LUCIA HALTY		
1.	Introducción	561
2.	Inteligencia artificial e Inteligencia artificial generativa en salud mental.....	563
2.1.	IA en salud mental	564
2.2.	Inteligencia artificial generativa: una nueva frontera	564
3.	Impacto positivo de la IAG en salud mental.....	565
4.	Impacto negativo de la IAG en salud mental	567
5.	Cómo fomentar un buen uso de la IAG en salud mental.....	569

5.1.	Regulación legal y ética: garantizar la innovación segura .	569
5.2.	Regulación emocional: la mejor defensa empieza en la infancia.....	571
6.	Conclusión: hacia un nuevo humanismo digital en salud mental.....	572
7.	Referencias bibliográficas.....	574
CAPÍTULO 22. ESPIRITUALIDAD Y ACOMPAÑAMIENTO		577
BERTA RUIZ - SARA LUMBRERAS		
1.	Introducción	577
2.	Espiritualidad y acompañamiento	578
3.	Irrupción de la IA en la espiritualidad y en el acompañamiento	581
4.	Riesgos de la IA en la espiritualidad y en el acompañamiento	583
5.	Recomendaciones éticas para su implementación y para su uso.....	585
6.	Conclusiones	586
7.	Referencias bibliográficas.....	587
CAPITULO 23. JOURNALISM AND DISINFORMATION A CRITICAL ANALYSIS		591
BOTOND FELEDY - MELINDA PINTÉR		
1.	Introduction.....	591
1.1.	Defining the Topic and Its Significance.....	591
1.2.	Research Aims and Questions	592
1.3.	Brief Methodological Note.....	593
2.	Historical and Theoretical Frameworks of Disinformation	593
2.1.	The Concept and Categories of Disinformation.....	593
2.2.	Historical Overview	593
2.3.	Communication-Theory and Psychological Foundations..	595
2.4.	New Forms in the Digital Age.....	596
3.	The Relationship Between Journalism and Disinformation.....	596
3.1.	The Press's Responsibility and Normative Role.....	596
3.2.	Journalistic Ethics and Disinformation Challenges	597
3.3.	Media Ownership Structures and Political Influence.....	598
3.4.	The News Race's Impact on Fact-Checking.....	598

4.	The Digital Revolution and the New Disinformation Ecosystem.....	599
4.1.	The Role of Social Media and Algorithms	599
4.2.	User-Generated Content and Citizen Journalism	600
4.3.	Memetics and Visual Disinformation	600
4.4.	Bots and Automated Disinformation Campaigns	601
5.	Case Studies	601
5.1.	Political Campaigns and Elections	601
5.2.	War and Crisis Situations	603
5.3.	Pandemics and Health Disinformation	603
5.4.	Local/Regional Examples (Hungary and Central Europe) .	604
6.	Fact-Checking and Defensive Strategies	606
6.1.	Fact-Checking Organizations and Methods	606
6.2.	Technological Solutions	606
6.3.	Media Education and Digital Literacy	607
6.4.	International Cooperation and Regulation	608
6.5.	EU initiatives	608
7.	Critical Approaches and Contested Questions	609
7.1.	Censorship vs. Freedom	609
7.2.	Risks of Political Abuse	609
7.3.	The Pluralism of <i>Reality</i>	610
8.	Conclusions.....	611
9.	Bibliographic references.....	613

CAPÍTULO 12.

BIOÉTICA, SALUD E INTELIGENCIA ARTIFICIAL

FEDERICO DE MONTALVO JÄÄSKELÄINEN
Universidad Pontificia Comillas

RAFAEL AMO USANOS
Universidad Pontificia Comillas

El buen médico trata la enfermedad;
el gran médico trata al paciente que tiene la enfermedad
William Osler

1. LA APLICACIÓN DE LA IA A LA RELACIÓN SANITARIA COMO MEJORA DE LA ASISTENCIA

La digitalización y el desarrollo de las nuevas tecnologías en el campo de la salud ofrece muchas oportunidades para la mejora en la asistencia y prevención de las enfermedades. Tal desarrollo viene, además, alentado por el gran número de datos que se recogen y tratan a diario en el ámbito de la práctica clínica, destacando la implantación de la historia clínica electrónica, como repositorio de toda la información asistencial con un identificador único por paciente que enlaza las diferentes bases de datos (Alcalde y Alfonso, 2019, p. 59).

Pero más allá del campo de la asistencia sanitaria más tradicional, que está ya dotada de potentes herramientas y sistemas tecnológicos, existen otros ámbitos de la realidad, no todos ellos estrictamente vinculados a la salud, en los que se han desarrollado sistemas inteligentes a través de *apps* y *wearables*, es decir, dispositivos electrónicos y aparatos *vestibles*, que se incorporan sobre alguna parte de nuestro cuerpo (ropa, gafas, pulseras, relojes ...) para producir datos sobre actividades, conductas o hábitos de vida¹. Los datos directamente derivados de la asistencia sanitaria, como

¹ Se calcula que existen más de 3,48 millones de aplicaciones móviles disponibles en la plataforma Google Play y 2,22 millones en el Apple App Store, de las que más de un millón están relacionadas con la salud, el estado físico, la alimentación y el bienestar general.

los de dichos, dispositivos crean un ecosistema perfecto para el desarrollo de algoritmos y el impulso de la IA.

El Comité Nacional de Bioética de Italia (2015), en su *Informe sobre los aspectos bioéticos de las aplicaciones móviles de salud* de 28 de mayo de 2015, destaca el papel que ya están ocupando las aplicaciones móviles o *apps* vinculadas directa o indirectamente al campo de la salud. También, el *Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información* (ONTSI) considera que las soluciones móviles en el entorno socio-sanitario han catapultado las posibilidades de mejora en la calidad del cuidado de la salud y la gestión sanitaria del paciente, proporcionando nuevas modalidades de asistencia y herramientas de diagnóstico. El objetivo de todo ello es mejorar el bienestar del paciente, erigiéndole como el foco de atención central del sistema sanitario, a la par que se favorece la gestación de modelos sostenibles y eficientes mediante la reducción de costes en la prestación de los servicios de salud. Su aplicación en el ámbito de la atención sanitaria posibilita actuaciones como la asistencia en situaciones de accidentes o emergencias de salud, la medición y seguimiento de parámetros físicos y fisiológicos del usuario en tiempo real y no real, así como el diagnóstico, tratamiento, prevención y promoción de la salud del paciente (ONTSI, 2016, p. 70). Y ello, sin necesidad de un contacto físico o proximidad.

Igualmente, otras tecnologías de IA como los *chatbots* pueden proporcionar un servicio 24/7, rápido y para muchas personas, resolviendo las tareas más tediosas y automáticas eficazmente. Entre éstos estarían, por un lado, los dirigidos a tareas automáticas como la resolución de consultas de los usuarios, el diagnóstico a partir de cuestionarios, etc. Son interacciones muy estructuradas y cerradas, reduciendo la posibilidad de error y facilitando su uso, planificación y gestión, de ahí que sean los más utilizados. Junto a ellos, el segundo tipo de *chatbots* son más complejos, ya que son capaces de interactuar con los usuarios y arrojar respuestas automáticas a las diferentes peticiones recibidas. Estos *chatbots* predictivos tienen algoritmos desarrollados con IA para detectar conductas, pensamientos o estados de ánimo gracias a la cantidad ingente de datos que recogen (*big data* y *data mining*). Los datos no generan conocimiento por sí solos; es su correcto procesamiento el que lo consigue al crear determinados patrones mediante la comparación (Baena, 2023, p. 77).

Así pues, la tecnología, y en particular la IA, puede ser una solución a la incertidumbre que rodea muchos aspectos de la práctica sanitaria, y especialmente, a cómo clasificar las condiciones de los pacientes (incertidumbre diagnóstica); por qué y cómo los pacientes desarrollan enfermedades (incertidumbre pato-fisiológica); qué tratamiento será más apropiado para ellos (incertidumbre terapéutica); o si se recuperarán con o sin un tratamiento específico (incertidumbre pronóstica) (Cabitza, Ciucci, Rasoni, 2019, p. 122).

2. APLICACIONES DE IA EN SALUD

Es difícil ofrecer un panorama detallado de las aplicaciones de la IA en salud sin que este quede desactualizado en cuestión de días. La velocidad con la que avanzan estas aplicaciones es vertiginosa, pero no se puede hacer una buena bioética sin unos buenos datos.

En general, y al margen de lo que hemos explicado antes con carácter general y a modo de introducción, la IA en salud tiene cinco grandes campos de aplicación: i.) tratamientos y cuidados personalizados; ii.) asistencia al diagnóstico médico; iii.) medicina preventiva y autocuidado; iv.) investigación biomédica y desarrollo de nuevos fármacos; v.) gestión optimizada de los hospitales y mejora de la calidad asistencial.

2.1. IA en diagnóstico

En los últimos años, los avances en la aplicación de IA al diagnóstico médico han dado resultados muy prometedores, fundamentalmente en dos campos: El del análisis de imágenes y el del análisis de datos clínicos.

Con respecto al análisis de imágenes, la IA en radiología ofrece sensibilidades equiparables o superiores a los especialistas en tareas de detección (cáncer de mama, pulmón, cerebro, etc.), con especificidades tan altas que varias herramientas han pasado del laboratorio a la clínica, obteniendo aprobaciones de agencias regulatorias de prestigio. Por ejemplo, IDx-DR para retinopatía diabética (Abràmoff, Lavin, Birch, et al.2018). La IA obtuvo una sensibilidad del 87.2% y especificidad del 90.7% en la detección de retinopatía. Es decir, identificó correctamente alrededor del 87% de pacientes con retinopatía significativa, con pocos falsos positivos ($\approx 9\%$ de falsos alarmas). La implementación de IDx-DR y herramientas similares podría aumentar significativamente la tasa de cribado de retinopatía diabética, especialmente en áreas con escasez de especialistas, manteniendo una alta sensibilidad para lesiones oculares tratables y aliviando la carga sobre los oftalmólogos.

En el campo del análisis de datos clínicos, existen modelos de IA entrenados con datos de historias clínicas electrónicas –incluyendo signos vitales, resultados de laboratorio y notas clínicas– para anticipar la aparición de sepsis o choque séptico horas antes de que clínicos lo reconozcan (LifeBridge Health News, 2024).

En este mismo campo, ya ha aparecido el uso de modelos de lenguaje natural que leen textos clínicos (notas de evolución, reportes, motivos de consulta) para extraer información relevante y apoyar decisiones diagnósticas. Los avances en modelos de lenguaje natural han permitido que las máquinas entiendan y resuman historiales médicos complejos ofreciendo diagnósticos, incluso en enfermedades raras (Mao, Huang, Jin, et al., 2025).

En conclusión, la IA aplicada al diagnóstico médico ha dejado de ser una promesa futurista para convertirse en una realidad emergente, respaldada por estudios recientes y aprobaciones oficiales, que está mejorando la detección precoz de enfermedades y apoyando la toma de decisiones clínicas en beneficio de los pacientes.

2.2. IA en tratamientos

La existencia de una inmensa cantidad de datos de salud ha permitido desarrollar algoritmos para analizar esos datos enormemente heterogéneos (genomas, historiales clínicos, biomarcadores, hábitos de vida, etc.) encontrando patrones que se escapan al ojo humano (Serrano et al., 2024). De este modo, la medicina avanza en la dirección de la personalización y la precisión maximizando la eficacia terapéutica y minimizando los efectos secundarios (Jamalinia y Weiskirchen, 2025). Especialmente llamativos son los avances en la oncología, aunque también se dan en otras especialidades. En concreto, los avances se producen i) en la predicción de eficacia farmacológica, ii) los modelos predictivos con datos genómicos y clínicos y, iii) el análisis automatizado de imágenes tumorales.

- i. Predicción de eficacia farmacológica. Uno de los límites que se presentaba a la oncología de precisión era la dificultad para interpretar el arsenal de datos genómicos que proceden de la secuenciación del perfil genético/molecular de un tumor. Los algoritmos de aprendizaje automático no son solo capaces de analizar rápidamente esos datos sino de integrarlos con los que proceden de otras fuentes, como la expresión génica, la proteómica, los datos clínicos, etc. Así, las herramientas de IA son capaces de refinar tratamientos oncológicos sugiriendo tratamientos óptimos para el perfil único de cada paciente, y ajustando la decisión al perfil individual del cáncer. En oncología la dosificación óptima puede variar entre pacientes. Así están creándose herramientas que permiten pasar del enfoque tradicional de dosis fija, basada en la dosis máxima tolerada, a modular la dosis específica para cada paciente según su tolerancia y respuesta, adaptándose a cambios en el estado clínico (Iasiak, Truong, Foo, et al., 2025). Estos avances en la oncología personalizada y de precisión permiten manejar mejor la resistencia tumoral a fármacos, identificando patrones de resistencia y sugiriendo alternativa, lo que permite predecir la eficacia de los tratamientos antes de su uso mediante modelos que enlazan biomarcadores del paciente con probabilidad de respuesta. Esta capacidad tiene un impacto enorme, no solo en la calidad de vida del paciente, sino en la reducción de costos para terapias tan caras como las inmunoterapias (Bendani, Boumajdi, Belyamani, et al., 2025).
- ii. Modelos predictivos con datos genómicos y clínicos. Los modelos predictivos impulsados por IA pueden combinar datos muy heterogéneos (desde genética

hasta hábitos de consumo) para estimar el desarrollo de una enfermedad o la posibilidad de recaída, la supervivencia estimada con distintos tratamientos o el riesgo de efectos adversos; todo ello con la peculiaridad de la personalización, es decir, para cada individuo. Como se ha indicado, la oncología es el lugar donde se producen resultados más prometedores. Guo et al. (2024) aplicaron aprendizaje automático interpretable para predecir el riesgo de metástasis a distancia en carcinoma de ovario de células claras (Mao, Shangguan, Huang, et al. 2025). Este modelo estima, en primer lugar, la probabilidad de metástasis en un paciente y, en segundo lugar, apunta los factores del paciente que impulsan ese riesgo, lo que permite un plan terapéutico personalizado. Más allá de la oncología, en cardiología y enfermedades hereditarias, se utilizan modelos de IA para estimar el riesgo individual de desarrollar ciertas patologías o complicaciones, combinando puntuaciones de riesgo genético (PRS), análisis de variantes patogénicas y datos del estilo de vida. Esto permite una medicina preventiva personalizada (Calvino, et al., 2025).

- iii. Análisis automatizado de imágenes tumorales. Tradicionalmente el análisis de imágenes de TAC, RMN, PET, etc., dependía de la pericia humana, sin embargo, la IA ha demostrado una capacidad extraordinaria para detectar patrones sutiles en imágenes tumorales. El análisis automatizado de imágenes tumorales mediante IA está proporcionando una nueva capa de información personalizada para la toma de decisiones. Características invisibles al ojo humano pueden ahora cuantificarse y correlacionarse con resultados clínicos, y aportar una información preciosa para la planificación de tratamientos. Patólogos computacionales utilizan redes neuronales convolucionales para analizar imágenes de láminas histológicas (tinción H&E, inmunohistoquímica, etc.) buscando patrones que correlacionan con comportamiento tumoral o respuesta a tratamientos. Por ejemplo, se ha desarrollado un clasificador profundo llamado PathoRiCH que analiza imágenes histopatológicas de cáncer de ovario para predecir si una paciente responderá a la quimioterapia basada en platino (Mao, Shangguan, Huang et al. 2025). Otra aplicación del análisis de imágenes por IA es la planificación radioterápica. Algoritmos de segmentación automática pueden delimitar con precisión milimétrica los volúmenes tumorales en scanner 3D, facilitando la definición de campos de radioterapia adaptados a la forma específica del tumor de cada paciente. Todo esto permite ajustar la dosis de radioterapia de forma personalizada.

2.3. Medicina preventiva y autocuidado

La IA está revolucionando la medicina preventiva y el autocuidado al analizar grandes cantidades de datos personales (historias clínicas, registros de wearables, apli-

caciones móviles) para anticipar riesgos de enfermedad, monitorizar la salud en tiempo real y promover intervenciones personalizadas antes de que surjan complicaciones.

Un ejemplo es el uso de *smartwatches* con sensores cardíacos y algoritmos de IA para identificar arritmias como la fibrilación auricular (FA) en etapas tempranas. Estudios recientes demuestran que relojes inteligentes con algoritmos avanzados pueden detectar episodios de FA en personas asintomáticas, facilitando un diagnóstico más temprano que con métodos tradicionales. Fitbit y Apple Watch han sido validados clínicamente para detectar ritmos cardíacos irregulares y pueden ser utilizados como herramientas de *screening* masivo (Lubitz, et al., 2022; Fiorina, et al., 2024).

Otra aplicación preventiva es la predicción de insuficiencia cardíaca y disfunción del corazón antes de que el paciente presente síntomas claros. En un estudio con casi 3.000 pacientes, una IA alcanzó un AUC de 0,85 con ~78% de sensibilidad y especificidad en la detección de disfunción ventricular, incluso identificando pacientes con ligera reducción de la función cardíaca que aún podrían beneficiarse de seguimiento (Eko Health Inc., 2025).

En el campo de la salud mental, la IA se emplea para predecir crisis psicológicas, detectar tempranamente trastornos y brindar apoyo personalizado a pacientes con depresión, ansiedad u otras condiciones. Un avance clave ha sido el análisis con IA de datos no estructurados –como publicaciones en redes sociales, texto libre o señales del teléfono– para detectar sutiles cambios de comportamiento o lenguaje que preceden a una crisis. Además de datos de texto, los sensores de móviles y *wearables* proveen información fisiológica y conductual valiosa. Algoritmos de aprendizaje profundo han combinado variables como ritmo cardíaco, variabilidad de la frecuencia, calidad del sueño, nivel de actividad física e interacción social (p. ej., uso del teléfono) para detectar cambios que suelen acompañar recaídas de depresión o brotes de trastorno bipolar (Mondal, 2025).

Los *wearables* también han mostrado utilidad para anticipar problemas agudos. Por ejemplo, durante la pandemia de COVID-19, proyectos académicos usaron IA sobre datos de pulseras y relojes (frecuencia cardíaca en reposo, sueño, etc.) para detectar signos tempranos de infección viral antes de la aparición de síntomas, permitiendo aislar casos presintomáticos y contener contagios. Del mismo modo, anillos inteligentes y relojes han logrado alertar de elevación inusual de pulso y temperatura corporal que precede a enfermedades como gripe o COVID, actuando como *alarmas* de salud preventiva que capacitan al usuario a buscar atención o descansar oportunamente.

En el ámbito del autocuidado del estilo de vida, las aplicaciones móviles con IA funcionan como coach digitales para prevenir enfermedades crónicas. Aplicaciones de actividad física y nutrición analizan los patrones del usuario y, mediante algoritmos adaptativos, brindan recomendaciones personalizadas para alcanzar objetivos saludables (pasos diarios, calidad del sueño, dieta equilibrada). Algunas incluso emplean IA

conversacional para motivar cambios conductuales –p.ej., un asistente virtual que recuerda al usuario moverse si ha estado sedentario mucho tiempo, o que sugiere ejercicios de relajación al detectar niveles de estrés elevados por el ritmo cardíaco. Estas funciones están respaldadas por evidencia: se ha visto que *apps* con mensajería inteligente y retroalimentación individual aumentan la adherencia a la actividad física, y mejoran parámetros como control glucémico en diabéticos o presión arterial en hipertensos, comparado con controles sin asistencia tecnológica (Babel, et al., 2021).

2.4. La IA en la investigación biomédica y el descubrimiento de medicamentos

Los algoritmos de IA están revolucionando el descubrimiento de fármacos al cribar enormes bases de datos químicas, moleculares y genómicas para identificar nuevas moléculas terapéuticas de forma más rápida y eficiente que los métodos tradicionales.

La IA está ayudando al descubrimiento de nuevos fármacos, el reposicionamiento de medicamentos y la optimización de ensayos clínicos. Uno de sus espectaculares primeros pasos fue la herramienta AlphaFold que realiza predicciones de estructura de las proteínas (Díaz-Holguín, 2024).

En el 2023 se han descubierto dos nuevos fármacos utilizando técnicas de *machine learning* (aprendizaje automático) y *deep learning* (aprendizaje profundo). Investigadores de MIT y McMaster entrenaron una red neuronal con datos de crecimiento bacteriano para predecir moléculas capaces de inhibir a *Acinetobacter baumannii*, una bacteria hospitalaria multirresistente y descubrir la *Abaucina* (Liu et al., 2023). Por otra parte, en menos de 30 meses, usando plataformas de *drug discovery* impulsadas por IA, se descubrió que la quinasa TNIK es un regulador clave de la fibrosis pulmonar idiopática, y se diseñó un inhibidor molecular selectivo llamado *rentosertib* para bloquearla (Xu, Ren, Wang et al., 2025).

La IA también se utiliza para el reposicionamiento de fármacos lo que optimiza mucho el trabajo ya realizado. Un caso emblemático es el de *baricitinib*, un fármaco aprobado para artritis reumatoide. Empleando su plataforma de IA con un extenso grafo bioinformático, la compañía BenevolentAI identificó en enero de 2020 que *baricitinib* podía ser útil contra el COVID-19 al interferir con mecanismos de entrada viral e inflamación del huésped (Richardson et al., 2022).

Tal es la potencia de estas aplicaciones que ya están surgiendo plataformas de IA diseñadas específicamente para el reposicionamiento a gran escala. Por ejemplo, el modelo TxGNN (*Therapeutics Graph Neural Network*) entrena una red neuronal gráfica sobre un masivo grafo biomédico (con 17.000 enfermedades, genes y medicamentos) para predecir qué fármacos existentes podrían tratar enfermedades que carecen de terapias disponibles (Huang, Chandak, Wang et al., 2024).

Por último, la IA ofrece herramientas para diseñar y ejecutar ensayos clínicos de forma más eficiente, mejorando varios aspectos clave: selección de pacientes, definición de protocolos y criterios, y monitoreo de datos en tiempo real para la seguridad y eficacia. Así agiliza y optimiza el proceso de llevar un fármaco al mercado mediante ensayos clínicos, que es notoriamente largo, costoso y con alta tasa de fracaso. La IA puede ayudar al diseño inteligente de los protocolos de ensayo haciéndolos más inclusivos y eficaces (Zhang, Zhang, Chen et al., 2023). También puede ayudar a la selección y reclutamiento de pacientes adecuados mediante sistemas automatizados de emparejamiento de pacientes con ensayos. Empleando procesamiento de lenguaje natural, estos sistemas leen tanto los protocolos de ensayo (criterios de elegibilidad) como las historias clínicas electrónicas de pacientes, para identificar candidatos idóneos de forma rápida.

2.5. Gestión optimizada de los hospitales y mejora de la calidad asistencial

El uso de IA en la gestión hospitalaria tiene grandes virtualidades. Es capaz de planificar la ocupación de camas y el flujo de pacientes con lo que se reduce el tiempo de espera la eficacia en la gestión (University College de Londres, 2022). De forma similar, algunos algoritmos anticipan picos de demanda en Urgencias para optimizar recursos y reducir tiempos de espera, lo que se ha traducido en mayor eficiencia operativa en hospitales europeos (Bellini et al., 2024).

La planificación quirúrgica y asignación de personal también se benefician de la IA. Modelos de *machine learning* (XGBoost, *random forests*, redes neuronales, etc.) han demostrado mejorar la predicción de la duración de cirugías, optimizar la programación de quirófanos y asignar los recursos postoperatorios (UCPA) de forma más eficiente (New Medical Economics, 2025).

La IA está liberando a los profesionales de tareas rutinarias en los hospitales. Un ejemplo es la codificación automática de diagnósticos y procedimientos a partir de textos clínicos. Otras tareas administrativas optimizadas con IA incluyen la programación de citas y agendas mediante asistentes virtuales que asignan citas de forma eficiente. Por ejemplo, chatbots sanitarios ya ayudan a gestionar citas, responder dudas comunes y hacer seguimiento de pacientes, aliviando la carga del personal administrativo (Comunidad de Madrid, 2024).

En Atención Primaria, la IA ofrece oportunidades para organizar la demanda asistencial de forma más proactiva y eficiente. Un frente importante es la priorización de pacientes y gestión de agendas. Mediante algoritmos de triaje inteligente, los centros de salud pueden clasificar solicitudes según urgencia y complejidad. Por ejemplo, ya existen *chatbots* y aplicaciones de autotriaje donde el paciente introduce sus síntomas y la IA sugiere si requiere atención inmediata, consulta en persona, teleconsulta

o autocuidado. Estos asistentes virtuales pueden aconsejar al paciente sobre el nivel de atención más adecuado (atención primaria vs. urgencias hospitalarias), evitando saturar los servicios de urgencias con casos leves (Naved y Luo, 2025).

La IA también facilita la identificación de pacientes en riesgo dentro de la población asignada a primaria, apoyando un cuidado preventivo. Analizando el historial clínico electrónico longitudinal (que en Atención Primaria suele contener años de datos de cada paciente), los algoritmos pueden prever la evolución de enfermedades crónicas o detectar quién tiene alta probabilidad de complicaciones.

Por último, la IA aporta valor en la predicción de brotes epidémicos y planificación poblacional. Al analizar grandes volúmenes de datos de salud pública y atenciones en primaria, los algoritmos pueden detectar tendencias anómalas o patrones tempranos que anuncien un brote. Por ejemplo, incrementos sutiles en consultas por síntomas respiratorios en varias clínicas podrían alertar de una ola de gripe antes de lo habitual. De hecho, combinando datos clínicos con otras fuentes (farmacia, búsquedas web, meteorología), ya se han logrado modelos predictivos de incidencia de gripe o COVID-19 con días de antelación.

En general, se puede decir que el uso de IA en la gestión mejora la eficacia y reduce costes. La IA también contribuye a la reducción de errores y eventos adversos. Al automatizar tareas propensas al fallo humano –como la codificación clínica o la vigilancia epidemiológica– se han observado mejoras en la seguridad. La satisfacción y experiencia del paciente también se ven impactadas positivamente. Al agilizar la atención y remover trabas burocráticas, los pacientes perciben un servicio más atento y humano. Un ejemplo ilustrativo es la herramienta *Scribe* de IA generativa desplegada en hospitales Quirónsalud: transcribe automáticamente la conversación médico-paciente y redacta el informe, de modo que el facultativo puede dedicarse a mirar al paciente y escucharlo en lugar de teclear en el ordenador.

3. LOS RETOS ÉTICOS DE LA IA EN EL ÁMBITO DE LA SALUD

La digitalización de la salud y la implementación de la IA en este ámbito también conlleva importantes retos éticos. Y así, el Comité de Bioética de España en su informe sobre el Borrador de Carta de Derechos Digitales de 27 enero de 2021, argumentaba que,

si bien la digitalización y avance de la tecnología son capaces de traer prosperidad y contribuir a la calidad de vida, ello solo es posible si se diseña e implementa de forma sensata. Para construir un mundo futuro que haga realidad estos propósitos, es necesario tomar en cuenta las consideraciones éticas y los valores morales compartidos. Estos deben ser interpretados como estímulos y oportunidades para la innovación, no como obstáculos y barreras. Y, a tales efectos, la proclamación de derechos y libertades y la implementación de las correspondientes garantías eficaces representa la

mejor manera de proteger tanto la dignidad como la libertad (Comité de Bioética de España, 2023).

La llegada de la IA difumina, en primer lugar, la diferencia entre aplicaciones de salud, que pertenecen a dispositivos médicos, y las aplicaciones que no pertenecen a los dispositivos médicos. Y es harto sencillo vincular a la salud una tecnología, aunque no haya sido específicamente desarrollada para tal fin, sin olvidar que ambos dispositivos están sujetos a diferentes controles sobre su seguridad y eficacia, y que muchos de los operadores de los dispositivos no médicos no son organismos científicos y están interesados en ingresar al mercado de la salud. Estos nuevos actores no están interesados en sujetar sus innovaciones a los estrictos requisitos del desarrollo de productos sanitarios, pero sí lo están en los datos que se extraen de los mismos. Varias de estas empresas mantienen su *distancia* de la salud en el desarrollo de las tecnologías, pero, una vez en el mercado por un procedimiento menos estricto al de los productos sanitarios, sí optan por vincularlos a la salud o, al menos, al bienestar. Y los usuarios acaban instalando la aplicación o utilizando el producto, no por su validación científica, sino por el rango de aprobación en la red expresado por los consumidores.

Y no debemos olvidar, tampoco, que la IA puede dar lugar a una nueva suerte de medicina defensiva, en la que el incremento de pruebas solo venga justificado por la necesidad que exige la propia tecnología de *alimentarla* con más y más datos de salud. El paradigma tecnológico puede alentar a los médicos a prescribir pruebas y análisis adicionales, no por su valor clínico, sino más bien por su utilidad para entrenar o probar sistemas de IA. Esto tiene implicaciones tanto en términos de aumento de los costos de la atención sanitaria como también de exposición de los pacientes a riesgos innecesarios de fuga de datos u otras violaciones de su privacidad. Cualquier generación de datos con valor clínico cuestionable o claramente motivada por su utilidad únicamente para la prueba o el desarrollo de sistemas de IA, aparentemente violaría el derecho a la privacidad del paciente.

Y, en relación con la medicina defensiva, es bueno recordar algo que ya ocurrió hace unas décadas con la implementación de la medicina basada en la evidencia y la protocolización de la asistencia. Siendo tal cambio de paradigma muy positivo para la salud de los pacientes y su correcto diagnóstico y tratamiento, también trajo consigo que la asistencia se basara exclusivamente en lo dispuesto por el correspondiente protocolo o guía, cuando, en determinados supuestos, no procedía aplicar el mismo a un paciente concreto. El protocolo es válido para un gran número de pacientes, pero no para todos, por lo que en ocasiones se acaba aplicando por el temor del médico a verse expuesto, ante un resultado no deseado, a una reclamación judicial frente a la que no pueda defenderse invocando el fiel y estricto seguimiento del protocolo. Los protocolos indican la forma de actuar en la mayoría de los casos, pero no siempre. Y esto mismo puede ocurrir con la IA ¿será capaz el médico de cuestionar las conclusiones

de la IA cuando considere que no se corresponden con el mejor diagnóstico o tratamiento según su juicio clínico? ¿Considerará el Tribunal de Justicia que se ha actuado correctamente cuando el médico desvía su decisión de lo que le sugiere la potente IA?

En definitiva, son muchos los retos éticos que plantea la incorporación de la IA al ámbito de la salud, sin perjuicio, obviamente, de los beneficios que conlleva un uso adecuado de la misma. Hemos apuntado ya dos de ellos, la difuminación de la diferencia entre productos sanitarios y no sanitarios y la nueva medicina defensiva basada en la IA. Ahora vamos a centrarnos en otros cuatro que creemos que tienen especial relevancia en el ámbito de la salud.

4. LA CORPOREIDAD COMO LÍMITE A LA IMPLEMENTACIÓN DE LA IA: MÁS ALLÁ DE LA MERA SUPERVISIÓN HUMANA

La digitalización de la relación médico-paciente y la implementación de la IA permiten que aquella pueda desarrollarse actualmente de cuatro modos diferentes:

- i. Mediante una relación presencial, como ha sido el modelo tradicional.
- ii. Mediante una relación personal, pero no presencial a través de la tecnología (correo electrónico, videoconferencia, SMS, etc.).
- iii. Mediante una relación digital, pero con supervisión médica. El médico interviene a posteriori evaluando la solución propuesta por la IA.
- iv. Mediante una relación puramente digital sin médico (una suerte de medicina sin médicos), sin presencia ni supervisión.

¿Cuál es el mejor modo para que se desarrolle la relación médico-paciente? La respuesta, si consultamos a los sujetos de la relación, dependerá de múltiples factores (edad, mayor o menor habilidad digital, mayor o menor disponibilidad de tiempo, nivel educativo, socioeconómico o cultural). También es cierto que sistematizar la relación en diferentes modos no significa que uno de ellos deba imponerse sobre los demás, excluyéndolos; sino que la tecnología permite ahora que dicha relación se desarrolle de diferentes maneras, aprovechando los beneficios y ventajas para el médico y el paciente de cada uno de ellos en diferentes momentos del proceso diagnóstico y terapéutico.

Y aquí radica precisamente uno de los principales retos de la IA en el ámbito de la salud, determinar cuándo puede implementarse plenamente, digitalizando por completo la relación y desapareciendo el carácter presencial de la misma, y cuándo debe preservarse indispensablemente éste, porque la relación médico-paciente es una relación de naturaleza singular no equiparable a otras relaciones profesionales. Quizás pudiera parecerse a la relación abogado-cliente, pero la diferencia es que los valores en juego en aquella son, incluso, de superior calado a los de ésta: vida, integridad física y psíquica, frente a libertad.

Por ello, a la hora de determinar cuál debe ser el modo que implementar en un ámbito o momento concreto de la relación, es importante recordar que las decisiones médicas no son decisiones técnicas ni científicas, sino también morales por los citados valores en conflicto y por el impacto que la enfermedad genera en el paciente.

El diagnóstico y el tratamiento del paciente por parte del médico no solo debe atender a lo que es mejor, desde la perspectiva estrictamente científico-médica para el paciente, sino que las decisiones también deben respaldar el bienestar moral o la autonomía del paciente como entidad con valor moral, en el sentido de que la decisión debe coincidir con los deseos y valores del paciente. Y ello, en el marco de una relación máquina-individuo, sin una relación conversacional no es fácil. El médico es un agente moral, pero la IA no. El médico es el canal por el que fluyen las intervenciones médicas hacia el paciente, en el sentido de que el médico debe estar de acuerdo con cada intervención realizada. En esta posición el médico tiene la obligación moral de actuar como guardián, salvaguardando el bienestar del paciente y reconociendo su complicidad en las intervenciones realizadas (Comité de Bioética de Castilla-La Mancha, 2023). Una exigencia de fidelidad que servirá de nexo entre la moralidad y los fines de la medicina y que se fundamentará en una indispensable confianza (Pellegrino y Thomasma, 1993, p. 65).

Se consulta a los médicos no simplemente como *enciclopedias de conocimiento*, como si de una IA generativa se tratara, sino más bien como expertos *confiables* capaces de realizar una evaluación subjetiva y comprender al paciente como una persona social, con una historia y unos valores (Emanuel y Emanuel, 1992). El médico no es solo un experto, sino también un aliado del paciente en el contexto de la enfermedad, porque ésta define singularmente la relación. No es una relación de proveedor con un consumidor, con un cliente; sino con alguien que, por la etimología de la palabra paciente, sufre, padece, se enfrenta a perder alguno de los valores más relevantes para el ser humano como son la integridad y, sobre todo, la vida.

La medicina, debemos recordar, es una ciencia aplicada, lo que significa que no solo busca explicar científicamente los fenómenos del mundo, sino que también es una disciplina prescriptiva. En otras palabras, busca intervenir en el mundo basándose en ciertas ideas sobre lo que es bueno o malo desde una perspectiva moral, no solo científica. La IA busca explicaciones causales para fenómenos empíricos, pero la relación médico-paciente no es solamente un fenómeno empírico. El cuerpo del paciente no es una estructura material pasiva que espera órdenes mentales, sino que participa activamente en una interacción inteligente y significativa con el entorno (Carel, 2016).

La presencia del médico en muchos momentos de la relación cobra una extraordinaria importancia por la propia naturaleza de la medicina. Es decir, el carácter presencial de la relación tiene un impacto en el proceso de diagnóstico y terapéutico, de

manera que puede afirmarse que, si se elimina por completo o, al menos, en aquellos momentos que es relevante para el proceso, se estaría ofreciendo una peor posibilidad de diagnóstico o curación. Se estaría incurriendo en una vulneración del principio de maleficencia, ya que se estaría perjudicando las opciones curativas del paciente.

Así, en un meta-análisis de seguimiento sobre la comunicación médico-paciente, se comprobó que la calidad de la comunicación durante la anamnesis y el tratamiento también afecta a los resultados del acto médico (p. ej., frecuencia de las visitas, salud emocional y resolución de los síntomas), así como que la manera en que un médico se comunica con un paciente (incluso mientras recopila información) influye en la frecuencia con la que un paciente volverá a consultar a ese mismo médico, si es que lo hace (Chipidza, Wallwork y Stern, 2015). En otro estudio se demostró que la adherencia al tratamiento mejoraba cuando junto al modelo de diálogo se utilizaban también las nuevas tecnologías, como modelo informativo. Esta combinación ofrecía mejores resultados que la información facilitada únicamente a través de las nuevas tecnologías y sin presencia del facultativo (Wu et al., 2022). Otro metaanálisis concluyó que la mejora de la relación médico-paciente supone beneficios equivalentes a los de algunos tratamientos farmacológicos (Kelley, 2014).

Como recuerda el Comité Nuffield del Reino Unido, si la atención de salud (en el contexto analizado por el informe, la mental) se vuelve cada vez más automatizada y sin contacto, es importante reflexionar sobre el impacto que la falta de contacto humano puede tener en quienes buscan atención. Para muchos usuarios de servicios, la terapia en persona, la participación comunitaria, el apoyo de pares y grupos y otras actividades que implican interacciones cara a cara pueden marcar la diferencia para mejorar los resultados clínicos. Los estudios iniciales que examinaron las experiencias de atención durante la COVID-19 demostraron que, si bien algunas personas se adaptaron rápidamente y apreciaron la flexibilidad que ofrecía la atención remota, otras experimentaron sentimientos exacerbados de soledad, aislamiento, sensación de desconexión de las comunidades y un deterioro general de la salud (mental) (Nuffield Council on Bioethics, 2022).

Porque, en palabras del mismo Comité, a algunos usuarios de servicios les puede resultar difícil establecer una relación de confianza con los profesionales de la salud sin un contacto cara a cara. Esto puede tener implicaciones para los resultados clínicos, dada la importancia de las relaciones terapéuticas para el proceso de recuperación. Y, a pesar de los recientes avances en la informática con elementos de afectividad, los sistemas automatizados todavía están muy lejos de comprender la experiencia subjetiva de una persona con respecto a la enfermedad mental. Tampoco se sabe si las máquinas algún día podrán replicar plenamente la riqueza de las emociones e interacciones humanas (Nuffield Council on Bioethics, 2022). Una pregunta para formularse, por tanto, es la del impacto que tiene en el bienestar y confianza del paciente el no ver a un médico (Swedish National Council on Medical Ethics, 2020).

Si a la hasta ahora relación médico-paciente, se suma un nuevo *actor* que sea percibido como una mera *caja negra* respecto de la que no se pueda identificar claramente su influencia en lo que está sucediendo, ello amenaza directamente la autonomía del paciente y la confianza en la relación médico-paciente. Porque el problema comienza con la cuestión de en quién se puede confiar o en quién no (Fritz, 2020, pp. 30-31).

Además, para que pueda generarse confianza y, sobre la base de ésta, ejercer adecuadamente la autonomía por parte del paciente, se exige que éste disponga de información adecuada. Tal información no necesariamente se obtiene sólo del médico, sino que también puede accederse a ella a través de otros medios, como son prospectos o la propia información que se contiene en internet. Sin embargo, la información que se obtiene en el marco de una relación personal médico-paciente, y no a través de la mera lectura, es más fácil que genere dicha confianza porque surge en el contexto del diálogo.

Y la información es considerada una de las piezas claves del sistema sanitario, y no sólo desde la perspectiva de la teoría de la voluntariedad del tratamiento, de manera que el paciente deba estar adecuadamente informado para decidir con verdadera autonomía (la información es, en este caso, el presupuesto de la autorización o rechazo del tratamiento) sobre la propuesta o propuestas de tratamiento que le ofrece el médico. La información es la herramienta principal de la prevención y lucha contra las enfermedades. La información garantiza que el paciente no sólo actúe para prevenir o diagnosticar en una fase inicial la enfermedad, al conocer sus riesgos particulares o los primeros síntomas, sino también permite que el paciente colabore de una manera más adecuada durante el tratamiento.

Y aunque el modelo tradicional se ha visto alterado por la aparición de nuevos agentes y productos de salud, las obligaciones de esta relación no han desaparecido. Antes al contrario, el desplazamiento de estas obligaciones hacia nuevos actores tecnológicos en la medicina es, precisamente, motivo de preocupación porque debe reflexionarse sobre cuál es la mejor manera de gobernar la introducción de la IA en la medicina. Dar primacía única a la autonomía del paciente, de manera que el médico quede reducido a un mero proveedor de servicios tecnológicos, hará que éste sea ya incapaz de dotar a la relación de las virtudes médicas y normas deontológicas de la práctica médica (Mittelstadt, 2021, pp. 42-43).

Se produce, así, una *descontextualización* de la condición del paciente, en la que el paciente pierde cierto control sobre cómo los médicos y cuidadores entienden su condición. Y todo esto plantea que los encuentros a través de los cuales tradicionalmente se desarrolla la confianza básica necesaria para una relación médico-paciente, pueden verse alterados por la mediación tecnológica. Las tecnologías que inhiben la comunicación de determinadas señales psicológicas y de las emociones, pueden impedir que el médico conozca la condición del paciente, socavando el establecimiento de una relación médico-paciente de confianza. Como mediador entre el médico y

el paciente, los sistemas de IA cambian las dependencias entre aquéllos al transferir parte de la atención continua del paciente a un sistema tecnológico. Esto puede aumentar la distancia entre los profesionales de la salud y los pacientes, lo que sugiere una pérdida de oportunidades para desarrollar una comprensión tácita de la salud y el bienestar del paciente (Mittelstadt, 2021, p. 52).

Todo lo que venimos comentando ha determinado que se exija la supervisión humana como garantía. Así, la supervisión humana aparece ya reconocida en la Ley de IA de la UE (Parlamento Europeo y Consejo de la Unión Europea, 2024). El artículo 14 dispone que «1. Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas», añadiendo que

2. El objetivo de la supervisión humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo, conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos establecidos en la presente sección.

Tal supervisión no ha de ser, además, meramente formal, sino, como ha señalado la Agencia Española de Protección de Datos (2023) en su Guía sobre la adecuación al RGPD de tratamientos que incorporan IA, «para que pueda considerarse que existe participación humana, la supervisión de la decisión ha de ser realizada por una persona competente y autorizada para modificar la decisión, y para ello ha de realizar una acción significativa y no simbólica». Y en lo que al ámbito sanitario se refiere, quien puede realizar dicha acción significativa, por ser persona competente y autorizada para modificar la decisión, es el profesional sanitario. A estos efectos, los términos en los que se expresa el artículo 3 de la Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, no pueden ser más elocuentes:

médico responsable: el profesional que tiene a su cargo coordinar la información y la asistencia sanitaria del paciente o del usuario, con el carácter de interlocutor principal del mismo en todo lo referente a su atención e información durante el proceso asistencial, sin perjuicio de las obligaciones de otros profesionales que participan en las actuaciones asistenciales.

E, igualmente, puede citarse el artículo 4 de la Ley 44/2003, de 21 de noviembre, de ordenación de las profesiones sanitarias que dispone que

4. Corresponde a todas las profesiones sanitarias participar activamente en proyectos que puedan beneficiar la salud y el bienestar de las personas en situaciones de salud y enfermedad, especialmente en el campo de la prevención de enfermedades, de la educación sanitaria, de la investigación y del intercambio de información con otros profesionales y con las autoridades sanitarias, para mejor garantía de dichas finalidades.

Añadiendo, a continuación, que

5. Los profesionales tendrán como guía de su actuación el servicio a la sociedad, el interés y salud del ciudadano a quien se le presta el servicio, el cumplimiento riguroso de las obligaciones deontológicas, determinadas por las propias profesiones conforme a la legislación vigente, y de los criterios de normo-praxis o, en su caso, los usos generales propios de su profesión».

Además,

«7. El ejercicio de las profesiones sanitarias se llevará a cabo con plena autonomía técnica y científica, sin más limitaciones que las establecidas en esta ley y por los demás principios y valores contenidos en el ordenamiento jurídico y deontológico.

Finalmente, la Carta de Derechos Digitales, aprobada por el Gobierno sin eficacia normativa, dispone en su apartado XXIII que

4. El empleo de sistemas digitales de asistencia al diagnóstico, y en particular de procesos basados en inteligencia artificial no limitará el derecho al libre criterio clínico del personal sanitario (Gobierno de España, 2021).

El Comité de Bioética de España (2023) en su Informe sobre el Borrador de Carta de Derechos Digitales, fue algo más lejos que lo que finalmente ha dispuesto la Carta. Así, señalaba el Comité que

junto a dicha garantía que encuentra su fundamento en la libertad profesional que consagran los artículos 35 y 36 de la Constitución, debiera recogerse otra, ya proclamada en el ámbito internacional, y que cobra especial importancia en el contexto actual de una salud digitalizada: que el uso de las tecnologías digitales sea un complemento, nunca como sustituto, al trato humano y la atención directa del sanitario con los pacientes. Y ello, tiene especial importancia en el ámbito de la asistencia sanitaria, ya que en la relación clínica hay elementos afectivos, de probada eficacia terapéutica, cuyo origen se halla en el carácter físico del encuentro personal en la atención sanitaria. La consideración ética de este elemento tiene su raíz en el principio de integridad que impide reducir a la persona a una simple razón autónoma. El ser humano no es reductible a datos. El ser humano tiene, por sí, elementos afectivos que el simple análisis cuantitativo de datos que se utiliza en algunas dimensiones de la llamada salud digital no agota, ni al propio ser humano ni a su concreto estado de su salud.

Para el Comité, «el principio de integridad por el cual se ha de considerar la dimensión corporal del ser humano obliga a un uso responsable de los elementos de la llamada Salud Digital que, en primer lugar, no interponga elementos entre el sanitario y el paciente que anulen la corporalidad y todo lo que a ella va ligado como la dimensión afectiva». Y recuerda que

la Comisión de Asuntos Jurídicos del Parlamento Europeo considera que el contacto humano es uno de los aspectos fundamentales del cuidado humano y cree que sustituir el factor humano por robots podría deshumanizar las prácticas de cuidado [...] Por lo tanto, sería importante proclamar el derecho a la salud digital y de acce-

so universal de la población a los dispositivos tecnológicos desarrollados con fines terapéuticos o asistenciales, pero sin detrimento alguno del contacto o trato humano, esencial en una relación como es la del sanitario y el paciente (Comité de Bioética de España, 2023).

La ya proclamada supervisión humana como garantía de la aplicación de la IA en la relación médico-paciente no es, pues, suficiente, porque supervisar no implica presencia del médico, sino participación en la decisión ofrecida por la IA, lo que puede tener lugar a posteriori, sin el paciente delante o sin haber conocido ni escuchado nunca al paciente.

La regulación de la relación médico-paciente en el entorno digital y los modos a través de los que puede desarrollarse la misma, no debe olvidar las notas esenciales de la relación médico-paciente que en muchos contextos del proceso de diagnóstico de la enfermedad y de su tratamiento exigen la relación interpersonal e, incluso, presencial de ambos actores. La relación médico-paciente es una asociación consensual en la que los pacientes buscan y aceptan la asistencia de un médico para gestionar su salud. Colaboran para lograr el más alto nivel de atención respetando la autonomía de los pacientes, comunicando y explicando las opciones y obteniendo el consentimiento informado. Los elementos clave de la relación son, por tanto, la comunicación efectiva y el respeto por las elecciones voluntarias, y su cumplimiento es independiente de la presencia de IA (Sutton, 2020). Las herramientas actuales basadas en IA carecen de la inteligencia contextual y emocional necesaria para tomar decisiones en circunstancias inciertas, de riesgo y emocionalmente tensas, y es que muchas decisiones no son simplemente una cuestión de lógica basada en la supervivencia (Lorenzini et al., 2023).

En palabras del Consejo Nacional de Medicina de Francia, por encima de todo, la relación médico-paciente es la clave para una buena atención médica práctica. Es el enlace esencial en la cadena de atención de salud, ya que sólo dicha relación puede garantizar,

- el respeto de la ética médica en el ejercicio de la medicina;
- un buen diagnóstico, empatía hacia los pacientes y la sensibilidad a sus circunstancias individuales;
- la elección del tratamiento adecuado a cada paciente;
- la rentabilidad del sistema de salud como un conjunto, preservando también los intereses del paciente;
- la salud del individuo y la del conjunto de sociedad (Conseil National de l'Ordre des Médecins, 2013).

O también, en palabras de la Asociación Médica Mundial en la Declaración de Córdoba sobre la relación médico-paciente: ésta es el núcleo fundamental de la práctica médica. Tiene un alcance universal y tiene como objetivo mejorar la salud y el

bienestar de las personas. Esto es posible gracias al intercambio de conocimientos, la toma de decisiones común, la autonomía del paciente y del médico, la ayuda, el consuelo y el compañerismo en una atmósfera de confianza. La confianza es un componente inherente de la relación que puede ser terapéutico en sí mismo. Y hoy en día se ve amenazada por una tecnificación de la medicina que, a veces, conduce a una visión mecanicista de la atención sanitaria, descuidando las consideraciones humanas. Y, por ello, la Asociación llama a preservar esta relación como núcleo fundamental de cualquier acción médica centrada en una persona, a defender la profesión médica y sus valores éticos, incluida la compasión, la competencia, el respeto mutuo y la autonomía profesional, y a apoyar a los pacientes (World Medical Association, 2020).

Y todo ello cobra pleno sentido si pensamos en el papel que la presencia humana tiene para el propio ánimo del paciente, siendo buena muestra de ello, del poder curativo del ánimo humano, los propios conceptos de placebo y nocebo. Este último viene a completar al de placebo, entendido éste, según lo define el Diccionario de la RAE, como aquella sustancia [acción de salud añadimos nosotros] que, careciendo por sí misma de acción terapéutica, produce algún efecto favorable en el enfermo, si este la recibe convencido de que esa sustancia posee realmente tal acción. Y el concepto de *efecto nocebo* se puede englobar entre aquellos efectos inespecíficos de un tratamiento que son perjudiciales, implicando, desde el punto de vista psicológico, la anticipación negativa en forma de ansiedad, miedo o repulsión al tratamiento. Esta expectativa negativa no solo determina la aparición de efectos adversos, sino que también puede disminuir la eficacia terapéutica del tratamiento o interferir en el curso de la enfermedad (Häuser, Hansen y Enck, 2012).

En la práctica diaria de los tratamientos, lo que realmente se ve es la suma de los efectos farmacológicos más el efecto placebo, las interacciones inconscientes entre el médico y el paciente. Y el efecto placebo es mucho más que el efecto de una tableta inerte. En él influyen las convicciones y el método de comunicación del terapeuta (González, Benedetto y Ramírez, 2008).

En definitiva, la proliferación e implementación de nuevas tecnologías en el marco de la relación médico-paciente debe reforzar el componente humano sobre el que se asienta la confianza (Morla, 2023, p. 98). En su defecto, se habrá perdido uno de los elementos que caracterizan la relación médico-paciente y el propio ejercicio de la libertad profesional del médico. No se trata ya de un mero cambio, sino de una mutación sustancial de aquella relación y de esta libertad.

En palabras de los Dicasterios para la Doctrina de la Fe y para la Cultura y la Educación, en su Nota sobre la relación entre la IA y la inteligencia humana, *Antiqua et Nova*, de 2025,

si la IA se utilizara no para mejorar, sino para sustituir por completo la relación entre pacientes y profesionales sanitarios, dejando que los primeros interactuasen con

una máquina en lugar de con un ser humano, se verificaría la reducción de una estructura relacional humana muy importante en un sistema centralizado, impersonal y desigual. En lugar de fomentar la solidaridad con los enfermos y los que sufren, estas aplicaciones de IA correrían el riesgo de agravar la soledad que suele acompañar a la enfermedad, sobre todo en el contexto de una cultura en la que “no se considera ya a las personas como un valor primario que hay que respetar y amparar”. Un tal uso de estos sistemas no sería conforme con el respeto de la dignidad de la persona y la solidaridad con los que sufren.

Y añade, a continuación,

la responsabilidad por el bienestar del paciente y las decisiones relacionadas que afectan a su vida constituyen el núcleo de la profesión sanitaria. Esta responsabilidad exige que el personal médico ejerza toda su capacidad e inteligencia para poner en práctica decisiones ponderadas y éticamente motivadas con respecto a las personas confiadas a su cuidado, respetando siempre la dignidad inviolable del paciente y el principio del consentimiento informado. En consecuencia, las decisiones relativas al tratamiento de los pacientes y la carga de responsabilidad asociada deben permanecer siempre en manos de las personas y nunca delegarse en la IA (Dicasterio..., 2025, nº 73-74).

Por ello, debería irse más allá de la mera exigencia de supervisión humana y garantizar el contacto humano, la presencia humana, la relación humana. El problema no está sólo en que el sistema funcione de manera autónoma, sin supervisión humana, sino que el sistema de IA sustituya o reemplace al ser humano en la relación con el otro en ámbitos en los que el juicio moral cobra especial relevancia, como es la relación médico-paciente.

Tal mínimo contacto o presencia humana sí es mencionado, aunque no proclamado expresamente como garantía, en el Reglamento UE sobre espacio europeo de datos de salud que señala que

podría ser poco ético informar a un paciente a través de un canal electrónico sobre un diagnóstico de una enfermedad incurable que probablemente sea terminal, en lugar de proporcionar esa información en primer lugar en una consulta con el paciente. Por lo tanto, debe ser posible retrasar la prestación del acceso a los datos de salud electrónicos personales en tales situaciones durante un período de tiempo limitado, por ejemplo, hasta el momento en que el profesional sanitario pueda explicar al paciente la situación (Unión Europea, 2024, Considerando 10).

Se aboga, por tanto, como demuestran diferentes estudios, por un modelo en el que las nuevas tecnologías asistan al médico, pero no le reemplacen. Los pacientes parecen aceptar mejor las nuevas tecnologías si actúan como asistentes de los médicos humanos en lugar de reemplazarlos. Incluso, se ha considerado que las nuevas tecnologías pueden facilitar una relación más humana entre el médico y el paciente en la medida que pueden darle más tiempo al médico para el diálogo (Sauerbrei, Kerasidou y Lucivero, 2023). La IA tiene el potencial de mantener la autonomía de los médicos, pero sólo con

la condición previa de una buena comunicación; de lo contrario, podrían convertirse en meros ejecutores pasivos, meros mensajeros, de las decisiones de la IA (Lorenzini et al., 2023). Las decisiones médicas no se basan únicamente en información clínica, sino que están entrelazadas con preferencias, valores, tolerancia al riesgo y muchos otros factores personales que deben sopesarse en el proceso de toma de decisiones (Lorenzini et al., 2023). Practicar la medicina es un arte y ninguna tecnología puede eliminar ese hecho. Cuando el médico se enfrenta a los pacientes, la interacción humana es vital. La satisfacción del paciente y la toma de decisiones compartida seguirán dependiendo en gran medida de este humanismo. La medicina es una profesión que todavía requiere compasión, tranquilidad y, lo más importante, empatía. Incluso con el advenimiento y la evolución continua de la IA y otros grandes modelos de lenguaje, la mejor manera de aprender y comunicar la empatía es mediante la enseñanza junto a la cama por parte de humanos, no de IA o *chatbots* (Glatter, Papadakos, & Shah, 2023). Además, IA no entiende los silencios, el ser humano sí, y aquéllos son tan importantes en la relación interpersonal como las palabras.

La empatía clínica no debe considerarse una simple postura humanística subjetiva. Además, una actitud empática no se opone a la objetividad de otras habilidades médicas y puede estudiarse con el rigor científico aplicado a otros ámbitos biológicos. Las expectativas específicas entre el paciente y su médico, cuando se cumplen y se atienden, reducen la incertidumbre y desempeñan un papel beneficioso y crucial en la curación. Un profesional atento que dedica más tiempo a conectar eficazmente con los pacientes es una enorme ventaja para la salud. Es beneficioso para los pacientes tener un médico que les dedica más tiempo y los escucha con atención (Decety, y Fotopoulou, 2015).

Por otro lado, no debemos olvidar que el médico es titular de la libertad de profesión, y que el fundamento de tal libertad es el de garantizar la prestación de un mejor servicio al paciente que se presume que se ha de lograr con mayor facilidad por quien es libre para decidir de acuerdo con su propio criterio técnico-científico, exento de toda presión externa. Tal libertad también se corresponde con una responsabilidad por parte de los profesionales de brindar la mejor atención a los pacientes. El estrecho vínculo entre la autonomía y la responsabilidad de los médicos se debe a la naturaleza de esta autonomía: se otorga a los médicos porque hacerlo proporciona un beneficio a la sociedad, y ese beneficio sería una buena atención (Lorenzini et al., 2023).

Los verdaderos fines de la medicina surgen como una respuesta a la experiencia humana universal de la enfermedad, que es intrínseca a la práctica de la medicina. Esta respuesta mueve a la necesidad de sanar, ayudar, cuidar y curar. La medicina comienza en la relación entre médico y paciente, que a su vez genera unos valores propios para mantenerse y reforzarse (como el mismo vínculo médico-paciente), a los cuales debería aferrarse para poder resistir los intentos de dominación o manipulación social y seguir un rumbo propio, además de sostener la integridad del

médico, independientemente de los valores sociales. La medicina se verá inevitablemente influida por los valores y fines de las sociedades de las cuales forme parte, pero esto no implica que sus propios valores puedan o deban someterse a los mismos.

Como señala el documento *Los fines de la medicina*, elaborado hace unos pocos años por una de las instituciones de la bioética más prestigiosas, *The Hastings Center* (2004), con la participación de quince expertos en bioética de diferentes partes del mundo, incluyendo la del español Diego Gracia; si bien la sociedad puede, en cierto modo, determinar el concepto y fines de la medicina, ello no significa que ésta esté mera y exclusivamente al albur de la decisión y evolución social. Si la medicina no posee un rumbo propio y unos valores fundamentales, la falta de recursos con los que resistir su invasión hará de ella una víctima fácil para la sociedad, como ocurrió durante el nazismo y el comunismo.

Para sus autores, si bien es indiscutible que las sociedades también han cambiado y solicitan de la medicina servicios que eran impensables e inauditos hace no muchos años, y ello, añadimos nosotros, debido en gran parte a la transformación de la medicina en un mero bien más de consumo y al desarrollo de las nuevas tecnologías, no todas las demandas sociales, o de los sectores más poderosos de la sociedad, merecen ser atendidas por igual. Ninguna ciencia, y la medicina no es una excepción, debe ser puesta al servicio de fines intrínsecamente malos. La medicina no puede emplearse en usos moralmente inaceptables, o, añadimos nuevamente nosotros, para fines que alteren las características sustanciales de la relación médico-paciente.

Pero es que, además, si la medicina responde a unos valores y principios ya consagrados a nivel universal, como expresa la Declaración Universal sobre Bioética y Derechos Humanos de la UNESCO (2006) aprobada en octubre de 2005, al señalar que «todos los seres humanos, sin distinción alguna, deberían disfrutar de las mismas normas éticas elevadas en la investigación relativa a la medicina y las ciencias de la vida», la medicina no puede quedar de la mano, en cuanto a su función y fines, de cada una de las sociedades en las que se desarrolla y ejerce la misma. El universalismo que se ha plasmado ya de manera ineludible en la bioética impide afirmar que la medicina dependa de la voluntad de cada sociedad, sobre todo, cuando afecta a valores y derechos de tan alto valor no solo constitucional, sino también universal como la dignidad, la vida y la integridad. Y menos aún podrá depender de lo que determine o decida una máquina. Tal libertad impide que terceros ajenos a la relación profesional influyan o intervengan en la toma de decisión por parte del profesional. Como tal libertad encuentra como principal destinatario al propio Estado que no sólo debe promover las condiciones necesarias para que dicha libertad quede garantizada, sino que además debe abstenerse de interferir en el ámbito de decisión del profesional.

Tal límite no solo afecta al legislador, a la hora de configurar los derechos y deberes del médico en su relación con el paciente, sino también a las empresas que, en

el ejercicio de su libertad, traten de desarrollar tecnologías que pretendan implementarse en el ámbito de la salud. Tales tecnologías no pueden desconocer cuáles son las funciones y garantías del médico en el ejercicio de su profesión, so pena de infringir la garantía institucional que supone la relación médico-paciente como expresión no solo de derechos de éste, sino también de la libertad profesional de aquél. La IA puede modificar la relación médico-paciente, pero no alterar sus elementos sustanciales y menos en perjuicio de sus dos protagonistas.

Por otro lado, para preservar la relación médico-paciente en la era de la IA, se ha propuesto aumentar también el enfoque en las habilidades sociales en el plan de estudios de medicina. Además de la importancia de la alfabetización en IA, la empatía y la compasión son habilidades que deben cultivarse o enseñarse a lo largo del plan de estudios y mantenerse activamente en el centro de la práctica médica, ya que los mejores médicos serán aquellos que comprendan cómo *se siente* el paciente. Y es que, además, el conocimiento y las habilidades técnicas ya no son del dominio exclusivo de la profesión médica, siendo aquél fácilmente accesible para el público. Por lo tanto, deben desarrollarse más otras competencias relevantes, como la capacidad de saber cuándo y cómo aplicar el conocimiento para ayudar mejor al paciente de manera compasiva, ya que lo verdaderamente importante es garantizar una atención adecuada en un futuro que va a depender cada vez más de herramientas de IA sin emociones (Sauerbrei, Kerasidou, Lucivero et al., 2023).

Como señala el Nuffield Council (2018) en su Nota sobre la IA y la asistencia sanitaria: «los humanos poseen atributos que los sistemas de IA podrían no poseer auténticamente, como la compasión. La práctica clínica a menudo implica juicios y habilidades complejos que la IA actualmente no puede replicar, como el conocimiento contextual y la capacidad de interpretar señales sociales. También existe un debate sobre si parte del conocimiento humano es tácito y no se puede enseñar».

La forma más eficaz de restaurar la empatía y la compasión como piedra angular de la comunicación del médico con los pacientes no es mediante modelos o un marco sugerido por la IA o los *chatbots*; antes al contrario, esto requiere un enfoque en la enseñanza y el diálogo de persona a persona. La educación en humanidades, ciencias sociales y la ciencia de la comunicación es tan vital como enseñar a los médicos anatomía y fisiología. En el pasado, la enseñanza junto a la cama en la facultad de medicina era un arte que se practicaba mediante la observación y la escucha cuidadosas, con especial atención a los ojos y el lenguaje corporal, transmitiendo formas no escritas y tácitas de expresar atención, preocupación y empatía por los pacientes. Esta interacción humana tiene implicaciones comprobadas para la satisfacción del paciente, la motivación y el cumplimiento de las recomendaciones de tratamiento (Glatter, Papadacos, & Shah, 2023).

En definitiva, si bien las nuevas tecnologías conllevan el riesgo de alterar una relación médico-paciente centrada en la persona e, incluso, desarrollar una nueva

forma de paternalismo basado en las presuntas verdades del algoritmo y a su falta de pluralidad de valores, también es cierto que pueden ser beneficiosas en la medida que pueden respaldar una toma de decisiones compartida al aumentar la autonomía del paciente y mejorar tal relación al suponer un ahorro de tiempo para el facultativo. La tecnología puede tener el potencial de recuperar la promesa de una atención médica más efectiva y eficiente y permitir a los médicos estar más atentos a sus pacientes. Si se emplean de forma eficaz, los avances en el análisis predictivo y la tecnología de la información podrían utilizarse para mejorar el flujo de trabajo clínico, mejorar nuestra comprensión del estado emocional y las necesidades humanas del paciente y conducir a una nueva era de humanismo en la medicina (Warraich, Califf y Krumholz, 2018).

Sin embargo, tales posibles ventajas solamente podrán hacerse realidad si se parte de la premisa de que las nuevas tecnologías son un mero apoyo en la toma de decisiones y no el actor principal de la relación, es decir, no reemplazan al médico y éste no solo se forma en las competencias y habilidades necesarias para el uso profesional de tales tecnologías, sino también en competencias y habilidades para que la relación médico-paciente siga considerándose una relación eminentemente humana. Pero mejorar la comunicación y las habilidades sociales de los médicos probablemente requerirá múltiples enfoques, como admitir estudiantes de medicina sobre la base de sus habilidades sociales y capacidad de empatía, capacitación temprana y continua en comunicación y construcción de relaciones (Nagy y Sisk, 2020). Y también que la participación de los médicos y pacientes desde las primeras etapas del desarrollo de las tecnologías médicas, para garantizar que estén centradas en la persona, sea real. Los tecnólogos e ingenieros que desarrollan tecnologías sanitarias deben tener experiencia en la prestación de atención sanitaria y en las características de la relación médico-paciente. Todo esto es necesario para desarrollar un compromiso común sobre un concepto de diseño en el que la tecnología y el cuidado humano no sean mutuamente excluyentes. La tecnología, si se aprovecha y desarrolla correctamente, puede enfatizar la presencia, en lugar de distanciar al médico de sus pacientes (Warraich, Califf y Krumholz, 2018).

Se necesitan, por tanto, más estudios empíricos, así como paneles de consenso de expertos, para desarrollar pautas generales que orienten a los pacientes y médicos sobre cómo utilizar el modo de comunicación apropiado según el contexto y cómo implementar la IA en su relación, preservando un contacto humano adecuado para el fin de la medicina que es el curar y cuidar.

5. JUSTICIA, SALUD E IA

Para Beauchamp y Childress el principio de justicia «se refiere a la distribución justa, equitativa y adecuada de beneficios y cargas, determinada por normas que es-

estructuran las condiciones de la cooperación social» (Beauchamp y Childress, 2024, p. 418). Es decir, que exige que los beneficios, riesgos y recursos se distribuyan de modo justo entre todas las personas. Ahora bien, como se ha indicado más arriba (cfr. Capítulo 9), la distribución justa será la que considere el bien común y la fraternidad universal, lo que implica la idea del destino universal de los bienes.

Con esta primera aclaración se delimita el contenido de este epígrafe. En primer lugar, y de forma negativa, no se trata aquí de la injusticia provocada por los sesgos de los que se va a hablar de inmediato, sino de la que procede del concepto de justicia o de errores –voluntario o involuntarios– de su implementación en los sistemas de IA en salud. En segundo lugar, de forma positiva, se trata aquí de las posibles injusticias que provengan de cualquier forma de discriminación y exclusión por el motivo que sea, y de las que tienen su origen en la falta de equidad.

Así las cosas, en el contexto de las aplicaciones de IA en salud pueden darse, al menos, cuatro tipos de injusticia que deberían tenerse en cuenta antes del desarrollo de estas aplicaciones. Las tres primeras se refieren a la discriminación y la cuarta a la falta de equidad.

La primera, es la injusticia que proviene de la pobreza de los datos de salud. Schmidth (2024), se refiere con ello a la capacidad reducida de ciertos grupos o países para beneficiarse de innovaciones en salud debido a su escasa representación en los datos utilizados por la IA. No se trata simplemente de un sesgo en los datos existentes, sino de una ausencia estructural que impide que estos colectivos sean representados en el entrenamiento de algoritmos, en las validaciones clínicas o en el acceso a las soluciones desarrolladas. Esta exclusión produce una doble forma de injusticia: epistémica, al no generar conocimiento sobre estas poblaciones; y distributiva, al negarles el acceso equitativo a los avances diagnósticos, terapéuticos o preventivos que la IA puede ofrecer, ya que quienes más necesitan las ventajas de la IA son los que quedan excluidos de ellas (Chinta, 2025).

La segunda, es aquella que se produce cuando los sistemas de IA que se aplican a la gestión sanitaria incorporan indicadores indirectos que reflejan desigualdades preexistentes y no son detectados por la falta de explicabilidad del algoritmo. Es una injusticia procedimental que causa exclusión de los beneficios de los sistemas de salud. Un ejemplo es los algoritmos utilizados en EEUU para asignar recursos, priorizar pacientes en listas de espera, predecir necesidades, basados en el gasto sanitario previo y no en la salud de los pacientes (Obermeyer, 2019). Es decir, en lugar de medir cuán enferma está una persona, el sistema asumía que quien había generado más gasto sanitario en el pasado era quien más necesitaba atención futura. Evidentemente la población pobre es la que menos gasto previo de salud realiza, por tanto, siguiendo al algoritmo, la que menos iba a necesitar. Se genera así un círculo vicioso de exclusión no solo por un sesgo de falta de datos, sino porque el criterio que se utiliza para programar el algoritmo ya lleva dentro de sí una injusticia.

La tercera es la injusticia por exclusión de las prioridades de investigación. Ocurre cuando los temas, poblaciones o enfermedades que se investigan están determinados por intereses económicos, tecnológicos o geopolíticos, en lugar de responder equitativamente a las necesidades sanitarias globales. Esta forma de injusticia se manifiesta, por ejemplo, cuando la IA se orienta preferentemente hacia enfermedades prevalentes en países ricos, descuidando problemas de salud que afectan desproporcionadamente más a poblaciones vulnerables o regiones empobrecidas, como las enfermedades tropicales olvidadas o la salud materna en contextos rurales. Como en los tres casos anteriores, no es un sesgo sino una injusticia previa a la IA que se incorpora a los sistemas de IA en investigación en salud por una opción determinada. Replica así una injusticia previa, algo contra lo que ya se ha pronunciado la OMS (2021) que solicitaba, entre otras cosas, que personas de poblaciones tradicionalmente excluidas se incorporen en el proceso de diseño de los algoritmos de IA para la investigación.

El cuarto tipo de injusticia es de naturaleza distinta. En este caso no es generada por discriminación sino por falta de equidad. Se origina cuando los beneficios derivados de la investigación con IA no llegan equitativamente a los distintos grupos sociales. Esto puede darse por dos motivos. El primero, muy cercano a los sesgos, ocurre cuando algoritmos de diagnóstico son entrenados con población europea cuya particularidad hace poco fiables sus resultados con población africana (Hao, 2020). El segundo, más propiamente referente al principio de justicia, ocurre de forma mucho más simple: cuando las poblaciones pobres no pueden adquirir estos sistemas de IA por su elevado precio, o, por ejemplo, habitantes con menos recursos no pueden acceder a dispositivos de autocuidado por el mismo motivo. Por esta razón, la Comisión Europea habla de la necesidad del *benefit sharing* (compartir beneficios) para que los beneficios de la IA no produzcan o agraven inequidades.

Se señala en *Antiqua et Nova* (Dicasterio ..., 2025, nº 74), ya citada antes, que el uso de la IA para determinar quién debe recibir tratamiento, basándose principalmente en criterios económicos o de eficacia, es un caso especialmente problemático de *paradigma tecnocrático* que debe rechazarse. De hecho, «optimizar los recursos significa usarlos de manera ética y solidaria y no penalizar a los más frágiles»; por no mencionar que, en este ámbito, dichos instrumentos están expuestos «a formas de prejuicio y discriminación. Los errores sistémicos pueden multiplicarse fácilmente, produciendo no sólo injusticias en casos concretos sino también, por efecto dominó, auténticas formas de desigualdad social» añadiendo que,

además, la integración de la IA en la atención sanitaria también plantea el riesgo de amplificar otras desigualdades ya existentes en el acceso a la atención. Dado que la atención sanitaria se centra cada vez más en la prevención y en los enfoques basados en el estilo de vida, puede darse el caso de que las soluciones impulsadas por la IA puedan involuntariamente favorecer a las poblaciones más acomodadas, que ya disfrutaban de un mayor acceso a los recursos médicos y a una nutrición de calidad. Esta tendencia corre el riesgo

de reforzar el modelo de una “medicina para ricos”, en la que las personas provistas de medios económicos se benefician de instrumentos avanzados de prevención y de información médica personalizada, mientras que a otros les cuesta acceder incluso a los servicios básicos. Por lo tanto, se necesitan marcos de gestión equitativos para garantizar que el uso de la IA en la atención sanitaria no agrave las desigualdades existentes, sino que esté al servicio del bien común (Dicasterio ..., 2025, nº 75).

6. SESGOS EN SALUD

Los sesgos son algo inherente al conocimiento. En los años setenta del siglo XX Kahneman y Tversky comenzaron a investigar los sesgos cognitivos y descubrieron, en el cerebro humano, las distorsiones y alteraciones que se producen al procesar información (Lazcoz, 2023, p. 183). Es decir, que el propio conocer –ya sea de la inteligencia humana o de la IA– implica limitaciones y desviaciones, porque la realidad es demasiado compleja como para ser conocida tal y cómo es, y porque el conocimiento, como la tecnología, no es neutro, como ha puesto de manifiesto la hermenéutica y la teoría del conocimiento.

Las taxonomías de tipos de sesgos son muy variadas. Algunos hablan de sesgos estadísticos o sociales (Cobo y Lloret, 2023, p. 51), de preexistentes, técnicos o emergentes (Lazcoz, 2023, p. 184). Ahora bien, por su naturaleza se puede hablar de cuatro tipos de sesgos: algorítmicos, en los datos, de implementación o éticos normativos. Los primeros son aquellos que, por el propio diseño del algoritmo, ya producen diferencias sistemáticas en el rendimiento; los segundos –que son la fuente principal de sesgo– son los que producen desigualdad y/o daño por la procedencia de los datos. Pueden ser porque se toman datos no representativos para alimentar los sistemas de IA o porque los datos que toman ya arrastran sesgos históricos de prejuicios. Los sesgos de implementación aparecen en el uso de la IA ya sea por la distribución de uso de la IA o por los errores en el uso. Por último, los sesgos éticos o normativos, es decir, los que proceden de la propia normativa, o autonormativa en el caso de la ética, para su aplicación.

En el campo de la ética de la IA, como afirma Lazcoz, los sesgos se han hecho muy «populares [...] por la repercusión que pueden tener desde un plano ético» (Lazcoz, 2023, p. 186). Las consecuencias éticas de los sesgos están relacionadas con dos de los principios bioéticos, la no-maleficencia y la justicia, ya que el sesgo puede causar un daño y/o una injusticia. A nivel algorético están relacionados con la robustez y solidez, pues se debe prever el daño causado por la IA con el fin de evitarlo, minimizarlo o indemnizarlo; y con la justicia, pues generan exclusión o inequidad.

El reto bioético que plantean los sesgos, traslada a los cinco campos de aplicaciones de la IA en salud: el diagnóstico, el tratamiento, el cuidado personal, la investigación y la gestión. Más allá del gran avance y mejora que supone el uso de la IA en

salud, en todos estos campos los sesgos pueden ser maleficentes o introducir injusticias. En un primer momento, nos referiremos a la relación entre sesgo y daño y, en un segundo momento, a la que se da entre sesgo e injusticia.

El común denominador de los sesgos que producen daño es que transforman la deficiencia técnica o la diferencia entre las personas en error y, cuando el error no se detecta y corrige, causa daño. Así, rompen la fiabilidad médica cuando afectan directamente al diagnóstico o al tratamiento y cuando introducen el error en el interior de la investigación.

Fundamentalmente son cuatro los daños que pueden causar los sesgos. El primero es que hacen perder precisión clínica en determinadas poblaciones ya que genera diferencias sistémicas en el diagnóstico y en las recomendaciones terapéuticas. Por ejemplo, alguna IA para la detección del cáncer de mama subdiagnostica a mujeres negras debido al sesgo algorítmico y de datos, provocando que el cáncer no se detecte a tiempo y conlleve menos supervivencia (Al-Bazzaz et al., 2024). En dermatología, por ejemplo, los sistemas de IA suelen estar entrenados casi exclusivamente con piel clara y tienen peor sensibilidad en la piel oscura (Gokhale et al., 2024). Algo similar ocurre en los oxímetros con PPG óptico pues producen lecturas inexactas en pacientes de piel oscura (Gokhale et al., 2024).

El segundo, son los errores del modelo, es decir, las desviaciones en la respuesta de un sistema de IA porque el ideal que se le propone como correcto a la hora del entrenamiento, no lo es. Por ejemplo, cuando se entrena a una IA para evaluar una cirugía y se sobreestima o subestima la habilidad técnica de un cirujano según su grupo de procedencia (Kiyasseh et al., 2023).

El tercer grupo de daños es el que procede del entrenamiento de la IA, en concreto de los datos que se le suministran. Algunos ejemplos pueden ilustrar la cuestión. El primero es una IA para la predicción de fractura ósea en mujeres que no calibra bien a la población negra, porque los datos introducidos no contemplan la subalimentación con historias clínicas de las mujeres de esa raza (Lo et al., 2023). El segundo es una IA para prever las necesidades de asistencia médica basándose en el gasto sanitario. La ausencia de datos de gasto por parte de algunos pacientes no indica la falta de necesidad sino la menor utilización del sistema (Obermeyer et al., 2019).

El cuarto grupo, procede del daño que puede causar la falta de formación de quien lo utiliza, al no calibrar correctamente ni comprender el problema del sesgo. Así, por ejemplo, puede darse el caso de médicos que confían ciegamente en la recomendación de la IA incluso cuando saben por su conocimiento que está errada (Khera et al., 2023).

Por lo que se refiere a los sesgos que pueden producir injusticias, es decir, pueden reproducir o amplificar desigualdades estructurales que ya existen en la sociedad con la consiguiente falta de equidad; hay que decir que estos sesgos surgen cuando se

ignora la diversidad humana real. Mayoritariamente proceden de los sesgos de datos y proceden de la desigual representación de la diversidad en las bases de datos que alimentan la IA. Valga, por ejemplo, que en las bases de datos genómicas, el 80 % de los estos proceden de europeos, con lo que los resultados de sistemas de IA que se alimentan de ellos son menos válidos para poblaciones africanas, asiáticas o indígenas (Celi et al., 2022; Hasanzadeh et al. 2025).

Si no se corrigen estos sesgos, las consecuencias son, en primer lugar, la consolidación, o incluso el aumento, de las desigualdades estructurales del pasado sin corregirlas. En segundo lugar, que se profundiza en un diseño excluyente de los futuros sistemas o aplicaciones de IA para salud que se apoyan sobre los anteriores, lo que genera un círculo vicioso excluyente para determinado grupo de personas. Lo mismo puede decirse, en tercer lugar, para la investigación en salud. La desigualdad se consolida cuando los datos de alimentación son ya sesgados, lo que conduce a unos resultados sesgados y excluyentes.

7. SOSTENIBILIDAD E IA

Estudios recientes señalan que la IA puede tener un impacto positivo en el logro de los ODS, especialmente en salud. Así según Saetra (2021) se han encontrado evidencias de que las aplicaciones de IA en salud han contribuido positivamente en el 69% de las sub-metas del ODS 3 (Objetivo 3: Garantizar una vida sana y promover el bienestar para todos en todas las edades), con solo un 8% de efectos negativos. Ahora bien, además estas aplicaciones también están relacionadas con la sostenibilidad ambiental, social y democrática.

De este modo, en primer lugar, nos referiremos a la contribución de las aplicaciones de IA en salud al ODS 3 y, en segundo lugar, abordaremos la contribución, o no, de estas aplicaciones a la sostenibilidad ambiental, social y democrática.

7.1. Aplicaciones en salud de la IA y su relación al ODS 3

En los cinco campos en los que se ha considerado el uso de las aplicaciones de IA en salud, se puede comprobar la magnitud de la contribución de estas aplicaciones al ODS 3, sin detrimento de que también han colaborado a otros ODS como el 4 (educación de calidad), el 5 (igualdad de género) y el 10 (reducción de desigualdades).

Los sistemas de IA para el diagnóstico contribuyen al ODS 3 al facilitar resultados tempranos y certeros que permiten adelantar el inicio de los tratamientos, lo que conlleva la reducción de mortalidad por enfermedades graves. De modo similar, el impulso a la medicina personalizada que la IA está imprimiendo en los avances médicos aumenta la calidad de los tratamientos y seguridad de los pacientes, lo que supone una clara contribución al ODS 3.

Especialmente significativa es la contribución de las aplicaciones de IA en salud en el campo de la salud pública y la prevención, pues contribuye a predecir brotes epidémicos y pronosticar la propagación de enfermedades, permitiendo a las autoridades sanitarias tomar medidas preventivas muy tempranas (Dankwa-Mullan, 2024). También, es igualmente reveladora la capacidad de los sistemas de IA aplicados a la gestión sanitaria ya que contribuye a hacer más con los mismos recursos, aumentando la eficacia operativa: reduciendo tiempos de espera, evitando desperdicios o, por ejemplo, mejorando el triage en las urgencias hospitalarias. Todo ello contribuye a mejorar la eficiencia y el bienestar y fortalecer los sistemas de salud (Palliyil, 2025).

7.2. Relación con la sostenibilidad ambiental, social y democrática

La sostenibilidad, comprendida de forma amplia, no solo se refiere a la protección del medio ambiente, sino también a la cohesión social y a la gobernanza democrática de las sociedades. En relación con las aplicaciones de IA en salud significa cómo asegurar que estas herramientas beneficien a todos los ciudadanos de forma justa, no dañen el medioambiente, se diseñen en atención a los valores democráticos y contribuyan a promover la democracia en las sociedades.

Por sostenibilidad medioambiental se entiende el impacto que el desarrollo, implementación y uso de las herramientas de IA en salud tienen sobre el medio ambiente, incluyendo el final de su vida útil y, con ello, la generación de recursos electrónicos.

La existencia y uso de estas aplicaciones tiene dentro de sí una contradicción, pues están diseñadas para salvar vidas pero, al mismo tiempo, pueden amenazarlas –especialmente a las más vulnerables– por el impacto medioambiental que tienen (Katirai, 2024).

Según señala Katirai existen tres puntos críticos en esta cuestión. El primero es el elevado consumo de energía y la consiguiente huella de carbono. Se estima en el 2025 la IA podrá consumir el 20 % de la producción eléctrica mundial y generar más de doscientos setenta mil kilos de CO₂. El segundo es la extracción de materiales necesarios para la construcción del hardware de la IA. Cada dispositivo de IA deja detrás de sí un reguero de contaminación en la extracción y procesamiento de tierras raras, litio, etc. El tercer punto crítico se refiere a la fase final de los dispositivos electrónicos que soportan la IA. Los sistemas de IA suelen actualizarse con mucha frecuencia generando una gran cantidad de basura electrónica que no siempre se recicla oportunamente, sino que se exporta a países con menor regulación.

A todo esto, hay que añadir que, dado que se ha mostrado que la crisis ambiental es una crisis socio-medioambiental y que quien más la sufre son los habitantes de los países pobres, se produce una gran injusticia: los ciudadanos que menos se benefician de la IA en salud –los pobres– son los que más sufren sus consecuencias.

La sostenibilidad social implica la cohesión social, es decir, que en la sociedad nadie se sienta excluido, lo que implica el hecho de compartir los beneficios. Las aplicaciones de IA en salud, como se ha indicado más arriba, juegan un gran papel en la cohesión social, pues su contribución al ODS 3, entre otros, supone una gran ayuda. Ahora bien, por el contrario, si estas aplicaciones no son diseñadas e implementadas con criterios éticos, como el de la beneficencia y la justicia, pueden alejar de la sociedad la ansiada cohesión (Saetra, 2021; Dankwa-Mullan, 2024).

En concreto, se ha de tener en cuenta que las aplicaciones de IA en salud pueden exacerbar las brechas sociales, generar la brecha digital en salud, replicar injusticias preexistentes o minar la confianza en los sistemas de salud y los médicos. Gouveia (2023) afirma que el desarrollo de sistemas de IA médica puede debilitar uno de los pilares fundamentales de la medicina: la confianza. Advierte que, cuando los diagnósticos son generados por algoritmos opacos, los pacientes ya no sabrán si están confiando en un médico humano con juicio clínico o en una máquina que opera sin contexto emocional o explicativo. Esto podría reducir la percepción de responsabilidad directa y afectar negativamente la relación médico-paciente.

Así las cosas, para que las aplicaciones de IA en salud sean sostenibles socialmente y contribuyan a la cohesión social debe contarse con la participación de personas de diversas procedencias en su diseño y desarrollo, tal y como afirma el documento Directrices éticas para una IA fiable (Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, 2019, nº 64).

Por último, la contribución de las aplicaciones de IA en salud a la sostenibilidad democrática implica que su gobernanza interna, su regulación y supervisión se haga en atención a los valores democráticos. Esto implica, fundamentalmente que los algoritmos sean explicables, la seguridad de los datos y la privacidad, la robustez y la solidez de los sistemas (Cfr. Capítulo 2), sin ello los algoritmos pueden no ser controlables, introduciendo en la sociedad incertidumbres que alejan el ideal de participación ciudadana.

No se puede olvidar en este punto el riesgo para la democracia que supone que el panorama tecnológico mundial esté dominado por un reducido grupo de empresas. La concentración de poder en torno a la IA puede amenazar la soberanía y la autodeterminación de los sistemas de salud más pequeños (Saetra, 2021).

8. REFERENCIAS BIBLIOGRÁFICAS

- Abràmoff, M. D., Lavin, P. T., Birch, M., et al. (2018). Pivotal trial of an autonomous AIbased diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*, 1, 39. <https://doi.org/10.1038/s41746-018-0040-6>
- Agencia Española de Protección de Datos (AEPD). (2023). *La adecuación al RGPD de tratamientos que incorporan inteligencia artificial*. <https://www.aepd.es/documento/adequacion-rgpd-ia.pdf>

- AlBazzaz, H., Janicijevic, M., & Strand, F. (2024). Reader bias in breast cancer screening related to cancer prevalence and artificial intelligence decision support—A reader study. *European Radiology*, 34(8), 5415–5424. <https://doi.org/10.1007/s00330-023-10514-5>
- Alcalde Bezhoid, G., & Alfonso Farnós, I. (2019). Utilización de tecnología Big Data en investigación clínica. *Revista Derecho General y Hospitalario*, núm. extraordinario, 2019.
- Alós, F., & PuigRibera, A. (2021). Uso de wearables y aplicaciones móviles (mHealth) para cambiar los estilos de vida desde la práctica clínica en Atención Primaria: una revisión narrativa. *Atención Primaria Práctica*, 3(S1).
- Baena, A. (2023). La roboética en salud. En VVAA, *Inteligencia artificial en salud. Retos éticos y científicos* (Cuadernos de la Fundación Víctor Grífols i Luca, n.º 63).
- Babel, A., Taneja, R., Mondello Malvestiti, F., Monaco, A., & Donde, S. (2021). Artificial Intelligence Solutions to Increase Medication Adherence in Patients With Noncommunicable Diseases. *Frontiers in Digital Health*, 3, 669869. <https://doi.org/10.3389/fdgth.2021.669869>
- Bellini, V., Ruso, M., Domenichetti, T., Panizzi, M., Allai, S., & Bignami, E. G. (2024, 26 de julio). Inteligencia artificial en la gestión de quirófanos [Artículo]. *SaludbyDíaz*. <https://saludbydiaz.com/2024/07/26/inteligencia-artificial-en-la-gestion-de-quirofanos/>
- Beauchamp, T. L., & Childress, J. F. (2025). *Principios de ética biomédica* (E. Valdés, Trad. e introd.; 1.ª ed.). Universidad Pontificia Comillas. (Cátedra de Bioética, 38)
- Cabitza, F., Batini, C., & Magni, M. (Eds.). (2019). *Organizing for the Digital World: IT for Individuals, Communities and Societies*. Springer.
- Calvino, G., Farro, J., Zampatti, S., Peconi, C., Megalizzi, D., Trastulli, G., Andreucci, S., Cascella, R., Strafella, C., Caltagirone, C., Grifalchi, F., & Giardina, E. (2025). From Genomics to AI: Revolutionizing Precision Medicine in Oncology. *Applied Sciences*, 15(12), 6578. <https://doi.org/10.3390/app15126578>
- Carel, H. (2016). *Phenomenology of illness*. Oxford University Press.
- Chinta, S. V., Wang, Z., Palikhe, A., Zhang, X., Kashif, A., Smith, M. A., Liu, J., & Zhang, W. (2025). AI driven healthcare: Fairness in AI healthcare: A survey. *PLOS Digital Health*, 4(5), e0000864. <https://doi.org/10.1371/journal.pdig.0000864>
- Chipidza, F. E., Wallwork, R. S., & Stern, T. A. (2015). Impact of the doctorpatient relationship. *Primary Care Companion for CNS Disorders*, 17(5).
- Comisión Europea. (s.f.). *Inteligencia artificial en la asistencia sanitaria*. EU Health Policy Platform. Recuperado 1 de julio de 2025, de https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_es
- Comité de Bioética de CastillaLa Mancha. (2023). *Informe sobre la relación clínicodigital: Aspectos éticos en el uso de las tecnologías digitales en la práctica clínica*. Servicio de Salud de CastillaLa Mancha. https://sanidad.castillalamancha.es/sites/sescam.castillalamancha.es/files/documentos/paginas/archivos/informe_sobre_la_relacion_clinico-digital_del_cb_de_clm_0.pdf

- Comité de Bioética de España. (2023). *Informe sobre el Borrador de la Carta de Derechos Digitales*. Instituto de Salud Carlos III. <https://comitedebioetica.isciii.es/wp-content/uploads/2023/10/Informe-CBE-sobre-el-Borrador-de-Carta-de-Derechos-Digitales.pdf>
- Conseil National de l'Ordre des Médecins. (2013). *The French Medical Council at the heart of the doctor/patient relationship* [Folleto institucional]. https://www.conseil-national.medecin.fr/sites/default/files/external-package/edition/1iqfpxu/cnom_plaquetteinstitutionelle_ev_0.pdf
- Cobo Cano, M., & Lloret Iglesias, L. (2023). *Inteligencia artificial y medicina*. Los Libros de la Catarata.
- DankwaMullan, I. (2024). Health equity and ethical considerations in using artificial intelligence in public health and medicine. *Preventing Chronic Disease*, 21, E30. https://www.cdc.gov/pcd/issues/2024/24_0245.htm
- DataFranco, J., & Berck, M. (2012). The nocebo effect: A clinician's guide. *Australian & New Zealand Journal of Psychiatry*, 7(6).
- Decety, J., & Fotopoulou, A. (2015). Why empathy has a beneficial impact on others in medicine: unifying theories. *Frontiers in Behavioral Neuroscience*, 8, 457. <https://doi.org/10.3389/fnbeh.2014.00457>
- DíazHolguín, A., Saarinen, M., Vo, D. D., Sturchio, A., Branzell, N., Cabeza de Vaca, I., Hu, H., MitjavilaDomènech, N., Lindqvist, A., Baranczewski, P., Millan, M. J., Yang, Y., Carlsson, J., & Svenningsson, P. (2024). AlphaFold accelerated discovery of psychotropic agonists targeting the trace amineassociated receptor 1. *Science Advances*, 10(32), eadn1524. <https://doi.org/10.1126/sciadv.adn1524>
- Dicasterio para la Doctrina de la Fe, & Dicasterio para la Cultura y la Educación. (2025, 28 de enero). *Antiqua et nova. Nota sobre la relación entre la inteligencia artificial y la inteligencia humana*. Vaticano. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_sp.html
- Eko Health Inc. (2025, 3 de marzo). Study: AI helps unlock earlier heart failure detection [Noticia]. *DAIC / Dlcardiology*. <https://www.dicardiology.com/content/study-ai-helps-unlock-earlier-heart-failure-detectio>
- Emanuel, E. J., & Emanuel, L. L. (1992). Four models of the physicianpatient relationship. *JAMA: The Journal of the American Medical Association*, 267.
- Fritz, A. (2020). The digital transformation in healthcare as a challenge for autonomy and trust in the physicianpatient interaction. *Dilemata*, 32, 17–36.
- Glatte, R., Papadakos, P., & Shah, Y. (2023, 23 de agosto). No AI can learn the art of medicine. *TIME*. <https://time.com/6306922/artificial-intelligence-medicine-doctors/>
- Gascón Abellán, M., González Carrasco, M. C., & Cantero Martínez, J. (2011). *Derecho sanitario y bioética. Cuestiones actuales*. Tirant lo Blanch.
- Gokhale, S., Daggubati, V., & Alexandrakis, G. (2024). Validation of a novel noninvasive technology to estimate blood oxygen saturation using green light: Observational study. *JMIR Biomedical Engineering*, 9(3), e46974. <https://doi.org/10.2196/46974>
- González Blasco, P., De Benedetto, M. A., & Ramírez Villaseñor, I. (2008). El arte de curar: el médico como placebo. *Atención Primaria*, 40(2), 93–95.

- Gouveia, S. S. (2023). *Podemos estar perante a última criação do ser humano: uma conversa sobre os dilemas da inteligência artificial com o filósofo Steven S. Gouveia*. Universidade do Porto. Recuperado de <https://repositorio-aberto.up.pt/bitstream/10216/156755/2/657034.pdf>
- Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial. (2019). Directrices éticas para una IA fiable. Comisión Europea. <https://digital-strategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai>
- Häuser, W., Hansen, E., & Enck, P. (2012). Nocebo phenomena in medicine: their relevance in everyday clinical practice. *Dtsch Arztebl Int*, 109(26).
- Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsowo, D. A., Azizi, Z., & White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *npj Digital Medicine*, 8(1), 154. <https://doi.org/10.1038/s41746-025-01503-7>
- Hastings Center. (2004). *Los fines de la medicina: El establecimiento de unas prioridades nuevas* (Trad. Fundación Víctor Grifols i Lucas). (Obra original publicada en 1996). Cuadernos de la Fundación Víctor Grifols i Lucas, (11).
- Huang, K., Chandak, P., Wang, Q., et al. (2024). A foundation model for cliniciancentered drug repurposing. *Nature Medicine*, 30, 3601–3613. <https://doi.org/10.1038/s41591-024-03233-x>
- Iasiak, A., Truong, A. T. L., Foo, N., et al. (2025). Personalized dose selection platform for patients with solid tumors in the PRECISE CURATE.AI feasibility trial. *npj Precision Oncology*, 9, 49. <https://doi.org/10.1038/s41698-025-00835-7>
- Katirai, A. (2024). The environmental costs of artificial intelligence for healthcare. *Asian Bioethics Review*, 16(3), 1–12. <https://doi.org/10.1007/s41649-024-00295-4>
- Kelley, J. M., et al. (2014). The influence of the patientclinician relationship on healthcare outcomes: A systematic review and metaanalysis of randomized controlled trials. *PLOS ONE*, 9(6), e101191.
- Khera, R., Simon, M. A., & Ross, J. S. (2023). Automation bias and assistive AI: Risk of harm from AI-driven clinical decision support. *JAMA*, 330(23), 2255–2257. <https://doi.org/10.1001/jama.2023.22557>
- Kiyasseh, D., Laca, J., Haque, T. F., et al. (2023). Human visual explanations mitigate bias in AI-based assessment of surgeon skills. *npj Digital Medicine*, 6, 54. <https://doi.org/10.1038/s41746-023-00766-2>
- Laurent Fiorina, P., Chemaly, P., Cellier, J., Ait Said, M., Coquard, C., Younsi, S., Salerno, F., Horvilleur, J., Lacotte, J., Manenti, V., Plesse, A., Henry, C., Lefebvre, B. (2024). Artificial intelligence-based electrocardiogram analysis improves atrial arrhythmia detection from a smartwatch electrocardiogram. *European Heart Journal - Digital Health*, 5(5), 535–541. <https://doi.org/10.1093/ehjdh/ztae047>
- Lazcoz Moratinos, G. (2023). Inteligencia artificial, sesgos y salud: Dilemas en torno a la propuesta de reglamento AIA. En R. Amo Usanos (Ed.), *Inteligencia artificial y bioética*. Universidad Pontificia Comillas.

- Lo, J. C., Chandra, M., Yang, W., Nath, S., Patel, M., Weintraub, M. R., Ettinger, B., et al. (2024). Challenges of fracture risk assessment in Asian and Black women. *The American Journal of Managed Care*, 30(3), 140–144. <https://doi.org/10.37765/ajmc.2024.89515>
- Lorenzini, G., et al. (2023). Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making. *Bioethics*.
- Lupton, D. (2016). *The quantified self*. Polity Press.
- Lubitz, S. A., Faranesh, A. Z., Selvaggi, C., Atlas, S. J., McManus, D. D., Singer, D. E., Pagoto, S., McConnell, M. V., Pantelopoulos, A., & Foulkes, A. S. (2022). Detection of atrial fibrillation in a large population using wearable devices: The Fitbit Heart Study. *Circulation*, 146(19), 1425–1435. <https://doi.org/10.1161/CIRCULATIONAHA.122.060291>
- Mao, X., Huang, Y., Jin, Y., et al. (2025). A phenotypebased AI pipeline outperforms human experts in differentially diagnosing rare diseases using EHRs. *npj Digital Medicine*, 8, 68. <https://doi.org/10.1038/s41746-025-01452-1>
- Mao, Y., Shangguan, D., Huang, Q., Xiao, L., Cao, D., Zhou, H., & Wang, Y. K. (2025). Emerging artificial intelligencedriven precision therapies in tumor drug resistance: recent advances, opportunities, and challenges. *Molecular Cancer*, 24(1), 123. <https://doi.org/10.1186/s12943-025-02321-x>
- Mohamad Jamalnia, Ralf Weiskirchen, et al. (2025). Advances in personalized medicine: translating genomic insights into targeted therapies for cancer treatment. *Annals of Translational Medicine*, 13(2), 18. <https://doi.org/10.21037/atm-25-34>
- Mondal, S. (2025, 27 de enero). How AI is detecting early warning signs of mental health disorders [Blog]. *AmerisourceCon*. <https://www.amerisourcecon.com/post/how-ai-is-detecting-early-warning-signs-of-mental-health-disorders>
- Morla González, M. (2023). *Medicina personalizada de precisión y derechos fundamentales*. Tirant lo Blanch.
- Nagy, M., & Sisk, B. (2020). How Will Artificial Intelligence Affect PatientClinician Relationships? *AMA Journal of Ethics*, 22(5).
- Naciones Unidas. (s.f.). Salud y bienestar: Objetivo de Desarrollo Sostenible 3. Recuperado 1 de julio de 2025, de <https://www.un.org/sustainabledevelopment/es/health/>
- New Medical Economics. (2025, abril). Inteligencia artificial en medicina: hacia un nuevo paradigma en la atención sanitaria [Artículo]. *New Medical Economics*. <https://www.newmedicaleconomics.es/inteligencia-artificial-en-medicina-hacia-un-nuevo-paradigma-en-la-atencion-sanitaria/>
- Nuffield Council on Bioethics. (2022). *The role of technology in mental healthcare* [Briefing note]. <https://www.nuffieldbioethics.org/publications/the-role-of-technology-in-mental-healthcare>
- Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información (ONTSI). (2011). *Los ciudadanos ante la esanidad: Opiniones y expectativas de los ciudadanos sobre el uso y aplicación de las TIC en el ámbito sanitario*. Ministerio de Industria, Turismo y Comercio. https://www.ontsi.es/sites/ontsi/files/los_ciudadanos_ante_la_e-sanidad.pdf

- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- Palliyil, V., Cai, M., Karam, H., et al. (2025). Artificial intelligence use cases adopted by people and their impact on achieving sustainable development goals: A systematic review [version 1; peer review: 1 approved with reservations]. *Open Research Europe*, 5, 117. <https://doi.org/10.12688/openreseurope.20023.1>
- Parlamento Europeo y Consejo de la Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea, L, 13 de junio de 2024. <http://data.europa.eu/eli/reg/2024/1689/oj>
- Pellegrino, E. D., & Thomasma, D. C. (1993). *The Virtues in Medical Practice*. Oxford University Press.
- Prainsack, B. (2020). The political economy of digital data: introduction to the special issue. *Policy Studies*, 41(5).
- Richardson, P. J., Robinson, B. W. S., Smith, D. P., & Stebbing, J. (2022). The AI-assisted identification and clinical efficacy of baricitinib in the treatment of COVID-19. *Vaccines*, 10(6), 951. <https://doi.org/10.3390/vaccines10060951>
- Romero Casabona, C. M. (2019). Revisión de las categorías jurídicas de la normativa europea ante la tecnología del big data aplicada a la salud. *Revista Derecho General y Hospitalario*, núm. extraordinario.
- Sauerbrei, A., Kerasidou, A., Lucivero, F., et al. (2023). The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. *BMC Medical Informatics and Decision Making*, 23, 73.
- Schmidt, J., Schutte, N. M., Buttigieg, S., Henschke, C., & Morley, J. (2024). Mapping the regulatory landscape for artificial intelligence in health within the European Union. *npj Digital Medicine*, 7, 229. <https://doi.org/10.1038/s41746-024-01221-6>
- Schweiger, A., & Parducci, A. (1981). Nocebo: the psychological induction of pain. *Pavlov Journal of Biological Science*, 16(3).
- Serrano, D. R., Luciano, F. C., Anaya, B. J., Ongoren, B., Kara, A., Molina, G., Ramirez, B. I., Sánchez-Guirales, S. A., Simon, J. A., Tomietto, G., Rapti, C., Ruiz, H. K., Rawat, S., Kumar, D., & Lalatsa, A. (2024). Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine. *Pharmaceutics*, 16(10), 1328. <https://doi.org/10.3390/pharmaceutics16101328>
- Sutton, R. T., et al. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3(1).
- Swedish National Council on Medical Ethics (SMER). (2020). *Artificial intelligence in healthcare: Ethical aspects* (Report No. 2020:2). <https://smer.se/wp-content/uploads/2020/06/smer-2020-2-in-brief-artificial-intelligence-in-healthcare.pdf>

- University College de Londres. (2022, 20 de septiembre). La inteligencia artificial puede reducir el tiempo de espera de un paciente para una cama de hospital [Artículo]. *RED Amgen*. <https://www.redamgen.com/actualidad/la-inteligencia-artificial-puede-reducir-el-tiempo-de-espera-de-un-paciente-para-una-cama>
- Villar Rojas, F. J. (1997). Medidas de racionalización; libertad de prescripción. *Derecho y Salud*, 5(2).
- VVAA. (2023). *Inteligencia artificial en salud. Retos éticos y científicos* (Cuadernos de la Fundación Víctor Grifols i Luca, n.º 63).
- Warraich, H. J., Califf, R. M., & Krumholz, H. M. (2018). The digital transformation of medicine can revitalize the patientclinician relationship. *npj Digital Medicine*, 1, 49.
- World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization. <https://www.who.int/publications/item/9789240029200>
- World Medical Association (WMA). (2020). WMA Declaration of Córdoba on patientphysician relationship. <https://www.wma.net/policies-post/wma-declaration-of-cordoba-on-patient-physician-relationship/>
- Wu, D., Lowry, P. B., Zhang, D., & Tao, Y. (2022). Patient trust in physicians matters—Understanding the role of a mobile patient education system and patientphysician communication in improving patient adherence behavior: field study. *Journal of Medical Internet Research*, 24(12).
- Xu, Z., Ren, F., Wang, P., et al. (2025). A generative AI discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial. *Nature Medicine*. <https://doi.org/10.1038/s41591-025-03743-2>
- Zhang, B., Zhang, L., Chen, Q., et al. (2023). Harnessing artificial intelligence to improve clinical trial design. *Communications Medicine*, 3, 191. <https://doi.org/10.1038/s43856-023-00425-3>

En un momento histórico en el que la inteligencia artificial ha dejado de ser una promesa técnica para convertirse en una infraestructura que atraviesa la vida social, este libro ofrece una brújula conceptual y práctica para orientarse en un territorio complejo. La IA se ha convertido en un entorno sociotécnico que reconfigura decisiones, instituciones, relaciones humanas y horizontes culturales. Comprenderla exige, por tanto, algo más que conocimiento técnico: requiere un marco ético capaz de discernir, ordenar y orientar la acción en contextos reales.

Esta obra responde a esa necesidad con una propuesta rigurosa y sistemática. Ofrece una arquitectura conceptual clara —fundamentos, núcleos temáticos y aplicaciones— que permite comprender la ética de la inteligencia artificial como una disciplina práctica, arraigada en los mecanismos reales de la tecnología y en los contextos concretos donde se despliega.

En sus páginas, el lector encontrará, en primer lugar, un sólido “suelo conceptual”: los presupuestos tecnológicos, filosóficos y jurídicos que condicionan cualquier evaluación ética. A partir de ahí, el libro identifica los grandes núcleos problemáticos que atraviesan la IA contemporánea, tratados no como categorías abstractas, sino como matrices interpretativas que permiten analizar cualquier sistema. Finalmente, el volumen desciende a los ámbitos de aplicación más mostrando cómo los dilemas éticos adoptan formas específicas cuando se encarnan en prácticas concretas.

El resultado es una obra interdisciplinar que dialoga con la ingeniería, la filosofía, la bioética, el derecho y las ciencias sociales, sin perder en ningún momento la claridad conceptual ni la orientación práctica. Dirigido a investigadores, profesionales y responsables institucionales, pero también a lectores que buscan comprender críticamente el presente, este libro parte de una convicción fundamental: la inteligencia artificial amplía el poder humano, pero no sustituye el juicio moral. Precisamente por ello, cuanto mayor es la capacidad técnica, más urgente se vuelve la exigencia ética.

**CÁTEDRA
DE BIOÉTICA**



9 791370 471590