ESTUDIOS

TOMANDO EL PULSO A LA RELACIÓN ENTRE INTELIGENCIA ARTIFICIAL, DERECHO Y ÉTICA

PABLO M. MELGAREJO CORDÓN
FRANCISCO HERRERA
DIRECTORES

Si quieres adquirir esta obra haz click aquí



UNIVERSIDAD DE GRANADA

DASCI
Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional

ARANZADI

© **Pablo M. Melgarejo Cordón y Francisco Herrera Triguero (Dirs) y autores**, 2026
© **ARANZADI LA LEY, S.A.U.**

ARANZADI LA LEY, S.A.U.
C/ Serrano Galvache 26
28033 Madrid
www.aranzadilaley.es
Atención al cliente: https://areacliente.aranzadilaley.es/publicaciones

Primera edición: Noviembre 2026

Depósito Legal: M-23558-2026
ISBN versión impresa: 978-84-1078-367-6
ISBN versión electrónica: 978-84-1078-368-3

Diseño, Preimpresión e Impresión: ARANZADI LA LEY, S.A.U.
Printed in Spain

© ARANZADI LA LEY, S.A.U. Todos los derechos reservados. A los efectos del art. 32 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba la Ley de Propiedad Intelectual, ARANZADI LA LEY, S.A.U., se opone expresamente a cualquier utilización del contenido de esta publicación sin su expresa autorización, lo cual incluye especialmente cualquier reproducción, modificación, registro, copia, explotación, distribución, comunicación, transmisión, envío, reutilización, publicación, tratamiento o cualquier otra utilización total o parcial en cualquier modo, medio o formato de esta publicación.
Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la Ley. Diríjase a **Cedro** (Centro Español de Derechos Reprográficos, **www.cedro.org**) si necesita fotocopiar o escanear algún fragmento de esta obra.
El editor y los autores no asumirán ningún tipo de responsabilidad que pueda derivarse frente a terceros como consecuencia de la utilización total o parcial de cualquier modo y en cualquier medio o formato de esta publicación (reproducción, modificación, registro, copia, explotación, distribución, comunicación pública, transformación, publicación, reutilización, etc.) que no haya sido expresa y previamente autorizada.
El editor y los autores no aceptarán responsabilidades por las posibles consecuencias ocasionadas a las personas naturales o jurídicas que actúen o dejen de actuar como resultado de alguna información contenida en esta publicación.
ARANZADI LA LEY no será responsable de las opiniones vertidas por los autores de los contenidos, así como en foros, chats, o cualesquiera otras herramientas de participación. Igualmente, ARANZADI LA LEY se exime de las posibles vulneraciones de derechos de propiedad intelectual y que sean imputables a dichos autores.
ARANZADI LA LEY queda eximida de cualquier responsabilidad por los daños y perjuicios de toda naturaleza que puedan deberse a la falta de veracidad, exactitud, exhaustividad y/o actualidad de los contenidos transmitidos, difundidos, almacenados, puestos a disposición o recibidos, obtenidos o a los que se haya accedido a través de sus PRODUCTOS. Ni tampoco por los Contenidos prestados u ofertados por terceras personas o entidades.
ARANZADI LA LEY se reserva el derecho de eliminación de aquellos contenidos que resulten inveraces, inexactos y contrarios a la ley, la moral, el orden público y las buenas costumbres.
Nota de la Editorial: El texto de las resoluciones judiciales contenido en las publicaciones y productos de **ARANZADI LA LEY, S.A.U.**, es suministrado por el Centro de Documentación Judicial del Consejo General del Poder Judicial (Cendoj), excepto aquellas que puntualmente nos han sido proporcionadas por parte de los gabinetes de comunicación de los órganos judiciales colegiados. El Cendoj es el único organismo legalmente facultado para la recopilación de dichas resoluciones. El tratamiento de los datos de carácter personal contenidos en dichas resoluciones es realizado directamente por el citado organismo, desde julio de 2003, con sus propios criterios en cumplimiento de la normativa vigente sobre el particular, siendo por tanto de su exclusiva responsabilidad cualquier error o incidencia en esta materia.

Si quieres adquirir esta obra haz click aquí

Índice General

Página

PREFACIO DE LOS EDITORES

PABLO M. MELGAREJO CORDÓN Y FRANCISCO HERRERA..... 35

PARTE I.
INTRODUCCIÓN GENERAL

INTELIGENCIA ARTIFICIAL, ÉTICA Y DERECHO: UNA PERSPECTIVA SOCIOTÉCNICA HACIA UNA CONFIANZA JUSTIFICADA

FRANCISCO HERRERA Y PABLO M. MELGAREJO CORDÓN..... 39

I. Introducción. Ética, Derecho y sistemas sociotécnicos: fundamentos para una inteligencia artificial confiable..... 39

II. Etica e Inteligencia Artificial: de los principios a la responsabilidad. 41

III. Derecho e Inteligencia Artificial: AI Act y su evolución mediante el digital omnibus on AI 45

3.1. Una arquitectura jurídica basada en los riesgos 45

3.2. Responsabilidades, transparencia y protección de las personas afectadas 46

3.3. La regulación de los modelos de inteligencia artificial de propósito general..... 47

3.4. El Digital Omnibus y la gobernanza institucional del AI Act..... 48

Si quieres adquirir esta obra haz click aquí

Página

3.5. Del cumplimiento jurídico a la inteligencia artificial confiable 50

IV. Inteligencia Artificial y sociedad 51

4.1. La inteligencia artificial como sistema sociotécnico 51

4.2. Dirección del cambio tecnológico, poder y progreso 53

4.3. Productividad, empleo y distribución de la renta 54

4.4. Desigualdad, daños sociotécnicos y grupos afectados 56

4.5. Concentración económica y control de las infraestructuras de IA 57

4.6. Sostenibilidad, cohesión social y finalidad de la innovación 58

4.7. Implicaciones para una inteligencia artificial confiable 59

V. Conclusiones 60

VI. Bibliografía 60

INTELIGENCIA ARTIFICIAL CONFIABLE

FRANCISCO HERRERA, OSCAR CORDÓN, MARÍA JOSÉ DEL JESUS, ENRIQUE HERRERA-VIEDMA Y MARCOS LÓPEZ DE PRADO 63

I. Introducción 63

II. Naturaleza y fundamentos de la Inteligencia Artificial confiable . 66

2.1. Confianza, confiabilidad y confianza justificada 66

2.2. Legalidad, ética y robustez 68

2.3. De los modelos a los sistemas sociotécnicos responsables 69

III. Requisitos y marcos de la Inteligencia Artificial confiable 71

3.1. Los siete requisitos de la inteligencia artificial confiable 71

3.2. Marcos de gestión de riesgos y estándares internacionales 73

IV. De los principios a la práctica: confiabilidad durante el ciclo de vida .. 75

4.1. Finalidad, contexto, riesgos e impactos 76

4.2. Diseño, desarrollo y evaluación del sistema 78

4.3. Despliegue, monitorización y evolución 80

V. Gobernanza, auditabilidad, rendición de cuentas y aseguramiento 82

5.1. Gobernanza y sistemas de IA responsables 82

Si quieres adquirir esta obra haz click aquí

Página

5.2. Auditabilidad y rendición de cuentas 83

5.3. Aseguramiento, auditoría y certificación 84

VI. Regulación, evidencia y aseguramiento de la IA confiable 86

VII. Desafíos emergentes y perspectivas de la Inteligencia Artificial confiable 89

VIII. Conclusiones 91

IX. Bibliografía 91

MÁS ALLÁ DE ABRIR LA CAJA NEGRA: EXPLICABILIDAD, JUSTIFICACIÓN Y COLABORACIÓN HUMANO-IA

FRANCISCO HERRERA 97

I. Introducción 97

II. ¿qué es la explicabilidad en Inteligencia Artificial? 100

III. Explicar, ¿para quién y para qué? 103

IV. Los límites de la explicabilidad 105

V. Más allá de explicar: contestabilidad, responsabilidad y gobernanza .. 108

VI. De las explicaciones estáticas a los co-explainers 112

VII. Conclusiones y retos futuros: XAI ante los grandes modelos de lenguaje 114

VIII. Bibliografía 117

PARTE II.
INTELIGENCIA ARTIFICIAL Y DERECHO

Marco constitucional, regulatorio
y de derechos fundamentales

LA INTERACCIÓN ENTRE IA Y CONSTITUCIÓN

FRANCISCO BALAGUER CALLEJÓN 121

I. La crisis de la constitución frente a la globalización y a la sociedad digital 121

Si quieres adquirir esta obra haz click aquí

Página

II. IA y constitución. Los problemas estructurales: el algoritmo sustituye al proceso 124

2.1. Las garantías formales del Estado constitucional 125

2.2. La transformación de la concepción del Derecho del Estado constitucional 126

III. IA y constitución. Los problemas coyunturales 128

3.1. Afectación a la democracia y los derechos 129

3.2. La IA generativa y su potencial de desinformación sistémica 136

IV. Conclusiones 139

LA LEY EUROPEA DE INTELIGENCIA ARTIFICIAL: UN HITO HUMANÍSTICO EUROPEO

IBÁN GARCÍA DEL BLANCO 141

I. Introducción: ¿Europa como isla? 141

II. El proceso regulador europeo: primeros pasos, la "soberanía digital europea" 144

III. El Reglamento Europeo dE Inteligencia Artificial, *iter* procedimental 147

IV. La Ley de Inteligencia Artificial: contenido y estructura 149

V. Otros contenidos destacables de la *AI ACT* 154

VI. Gobernanza e implementación 155

VII. Régimen sancionador 156

VIII. Conclusión 157

EL REGLAMENTO EUROPEO DE IA: DE LA TEORÍA A LA PRÁCTICA. EL CAMINO DE SU IMPLEMENTACIÓN

MIGUEL VALLE DEL OLMO 159

I. Introducción 159

Si quieres adquirir esta obra haz click aquí

Página

II. Origen y encaje jurídico del Reglamento de IA-seguridad de productos 160

2.1. Ejemplos para comprender la naturaleza de los riesgos 160

2.2. El nuevo marco legislativo y el viejo marco legislativo de seguridad de productos 164

III. Estructuras de gobernanza 166

3.1. El enfoque de la gobernanza nacional 166

3.2. Gobernanza Europea. AI Office y AI Board 167

3.3. Nexo entre autoridades de supervisión: coordinación vertical 169

IV. Cumplimiento de las obligaciones: responsabilidades de diferentes operadores 170

V. Códigos de práctica de los modelos de IA de propósito general 174

VI. Prohibiciones: breve descripción de las recientes guías de la comisión 175

VII. Sandboxes regulatorios 178

VIII. Otros elementos: alfabetización de IA, clasificación del alto riesgo 179

IX. Conclusiones 179

LEVANTAR EL VELO ALGORÍTMICO: LA TUTELA JUDICIAL EFECTIVA FRENTE A LAS DECISIONES ADOPTADAS O ASISTIDAS POR INTELIGENCIA ARTIFICIAL

PABLO M. MELGAREJO CORDÓN Y BEGOÑA ÁLVAREZ CIVANTOS .. 181

I. Introducción 181

II. La tutela judicial efectiva como garantía frente al poder algorítmico 183

2.1. Del juez frente al poder público al juez frente al algoritmo 183

2.2. La inteligencia artificial como nuevo centro de imputación funcional, sin personalidad jurídica 184

2.3. El artículo 47 CDFUE y el artículo 24 CE: un doble eje de protección tras el caso Bosco 186

Si quieres adquirir esta obra haz click aquí

Página

III. El derecho a comprender la decisión automatizada 187

3.1. Transparencia, trazabilidad y explicabilidad: delimitación conceptual 187

3.2. Explicabilidad técnica y explicabilidad procesal: una distinción necesaria 188

3.3. El derecho a conocer la lógica de la decisión: del artículo 22 RGPD al acceso al código fuente 190

3.4. La motivación como presupuesto del control jurisdiccional 192

IV. El acceso a la justicia en la era de la Inteligencia Artificial 194

4.1. ¿Contra quién se demanda cuando decide un algoritmo? El problema de la legitimación pasiva 194

4.2. La cadena de responsabilidad: imputación por diseño, por implementación y por decisión final 196

4.3. La pluralidad de operadores y la fragmentación de responsabilidades 197

4.4. Del juez asistido al juez "aumentado": el control judicial frente a sistemas de alta complejidad 198

V. La prueba frente a la opacidad algorítmica 200

5.1. La asimetría informativa entre ciudadano y proveedor 200

5.2. La carga de la prueba en los litigios derivados de la IA: el contraejemplo de State v. Loomis 200

5.3. Hacia la atenuación de la carga de la prueba: la presunción de causalidad algorítmica 201

5.4. Auditorías algorítmicas, pericial informática y acceso a la evidencia digital 202

VI. ¿Es suficiente la tutela judicial tradicional? Hacia nuevos mecanismos de protección 202

6.1. Los límites del proceso civil clásico frente al litigio algorítmico 202

6.2. La orden de exhibición algorítmica como medida cautelar autónoma. 203

6.3. Un registro judicial de sistemas de IA: inscripción, publicidad y control previo 203

Si quieres adquirir esta obra haz click aquí

Página

6.4. La responsabilidad procesal escalonada según el nivel de riesgo del sistema 204

6.5. Balance crítico y propuestas de lege ferenda 205

VII. Conclusiones 206

VIII. Bibliografía 207

AI ACT, DIGNIDAD HUMANA Y CONFIABILIDAD SOCIOTÉCNICA: CUMPLIMIENTO, EVIDENCIA Y ASESORÍA LEGAL PARA UNA IA CONFIABLE

BEGOÑA ÁLVAREZ Y FRANCISCO HERRERA 209

I. Introducción 209

II. De la IA confiable a la confiabilidad sociotécnica 212

III. El AI Act como infraestructura jurídica de confianza 214

IV. Cumplimiento, evidencia y asesoría legal como condiciones operativas 215

V. Dignidad humana, derechos fundamentales y confiabilidad sociotécnica 217

VI. Inferencia razonada, contestación y reconstrucción de la decisión asistida por IA 219

VII. Preguntas abiertas y orientaciones futuras para el ciclo de vida de una IA confiable 224

VIII. Conclusiones 226

IX. Bibliografía 227

LA APLICACIÓN EXTRATERRITORIAL DEL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL

ALFONSO ORTEGA GIMÉNEZ 231

I. Planteamiento 231

Si quieres adquirir esta obra haz click aquí

Página

II. La aplicación extraterritorial del Reglamento Europeo de Inteligencia Artificial, a imagen y semejanza del Reglamento General de Protección de Datos de la Unión Europea 238

III. Reflexiones finales 244

IV. Bibliografía consultada 247

VÍAS DE RECLAMACIÓN DE LOS PARTICULARES EN EL MARCO DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

RICARDO RUEDA VALDIVIA 253

I. Introducción 253

II. Reclamación de los particulares por infracción del reglamento de Inteligencia Artificial 259

2.1. Inexistencia de una única autoridad competente para conocer de las posibles reclamaciones: distribución de competencias entre autoridades nacionales y europeas 259

2.2. Criterios para la determinación del Estado miembro al que acudir para la formulación de reclamaciones cuyo conocimiento corresponda a autoridades nacionales 261

2.3. Procedimientos sancionadores y sanciones 265

III. Reclamación de los particulares por daños derivados del uso de la IA 270

3.1. Responsabilidad e IA: aproximación al estado de la cuestión en la UE . 270

3.2. Reclamación de responsabilidad en vía judicial 275

3.2.1. Determinación del órgano jurisdiccional competente para conocer de la reclamación de responsabilidad 275

3.2.2. Determinación de la ley aplicable a la responsabilidad 282

3.3. Reclamación de responsabilidad en vía extrajudicial 290

IV. Bibliografía 293

Si quieres adquirir esta obra haz click aquí

Página

INTELIGENCIA ARTIFICIAL Y DERECHOS FUNDAMENTALES EN INTERNET: LÍMITES PARA LA RESPONSABILIDAD PENAL

JESÚS IVÁN MORA GONZÁLEZ 297

I. Inteligencia Artificial y Derechos Fundamentales en Internet: reflexión sobre la necesidad de la sanción penal 297

II. Requerimiento de conocimiento específico como garantía para la delimitación de la responsabilidad penal: alcance de la posición de garante 300

III. La participación omisiva y la posición de garante: necesidad de una interpretación garantista de la normativa comunitaria 304

3.1. Requisitos formales para la generación de un conocimiento específico de presunta infracción 305

3.2. Generación del conocimiento específico de la infracción: necesidad del derecho de defensa 306

3.3. Incumplimiento del deber de impedir la continuidad del acto ilícito: alcance de la responsabilidad del intermediario 307

IV. Conclusiones 308

V. Bibliografía 309

Sectores jurídicos específicos

INTELIGENCIA ARTIFICIAL Y DERECHO CIVIL: ESPECIAL REFERENCIA AL CONSUMO

GUILLERMO OROZCO PARDO Y MARGARITA OROZCO GONZÁLEZ. 315

I. Introducción. El derecho civil en la sociedad digital 315

II. Una reflexión: IA y fuentes del derecho 319

III. Los derechos de las personas en la sociedad digital 321

IV. Regimen jurídico de la IA 324

V. El estatuto jurídico de las "personas electrónicas" 328

VI. La información es el "precio" 330

VII. IA, consumo, mercado digital y contratos inteligentes 331

Si quieres adquirir esta obra haz click aquí

Página

VIII. Aplicación de la IA en el ámbito de los contratos 337

IX. En breve: la *neuroprivacidad* como la nueva frontera y la necesaria "*reconceptualización* de las categorías jurídicas asentadas 341

LA INSTITUCIÓN REGISTRAL Y LA INTELIGENCIA ARTIFICIAL: EL TRILEMA DE EHRENBERG EN LA ERA DIGITAL

ANA GÓMEZ ADEVA 347

I. Introducción 347

II. Aplicabilidad e incidencia del Reglamento de Inteligencia Artificial en los registros 348

III. Delimitación de las tareas del procedimiento registral según impliquen funciones decisorias 350

3.1. Procedimiento registral ordinario 350

3.1.1. Presentación 350

3.1.2. Calificación 354

3.1.3. Redacción del asiento 356

3.2. Publicidad registral 358

3.3. Medios Adecuados de Solución de Controversias (MASC) 359

3.4. Colaboración con las Administraciones Públicas 361

IV. Otras funciones en ciernes 363

4.1. Tokenización de activos 363

4.2. Especial referencia al Registro de Bienes Muebles 365

V. Conclusión 366

VI. Bibliografía 368

EL PODER DE DECISIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN TRIBUTARIA

MARÍA BERNAD FUREST 371

I. Introducción 371

Si quieres adquirir esta obra haz click aquí

Página

II. Concepto de Inteligencia Artificial 371

III. La Inteligencia Artificial en la Agencia Estatal de la Administración Tributaria 375

IV. La Inteligencia Artificial y la automatización de sus actos 378

V. Inteligencia Artificial decisional y *Big Data* 381

VI. La Inteligencia Artificial y los derechos de los contribuyentes 384

VII. Conclusiones 386

VIII. Bibliografía 388

¿HACIA UN CONVENIO SOBRE LA DELINCUENCIA?

ALFONSO PERALTA GUTIÉRREZ 391

I. Introducción 391

II. Delitos contra la vida y la salud de las personas 396

2.1. Homicidio y asesinato 396

III. Delitos contra la libertad e indemnidad sexual 401

IV. Ciberataques, fraudes y estafas 405

V. Otros delitos y cuestiones 409

VI. Conclusiones 414

VII. Bibliografía 415

INTELIGENCIA ARTIFICIAL Y DERECHOS EN EL ÁMBITO DEL TRABAJO: DE LO INDIVIDUAL A LO COLECTIVO

OLIMPIA MOLINA HERMOSILLA Y ROSA M.ª GONZÁLEZ DE PATTO . 417

I. Planteamiento general 417

II. Hacia una visión integradora entre Inteligencia Artificial y derechos fundamentales en los entornos laborales 418

2.1. Nuevas necesidades de protección, nuevas categorías jurídicas 421

2.2. La Carta de Derechos Digitales en España 422

Si quieres adquirir esta obra haz click aquí

Página

III. Gobernanza colectiva de los entornos laborales digitales 423

IV. Tratamiento de los derechos digitales en sede de negociación colectiva 427

4.1. El RIA y la función de la negociación colectiva 427

4.2. Oportunidades y retos de la negociación colectiva en la protección de los derechos de los trabajadores frente a los riesgos de la IA 429

4.3. El estado de la cuestión en la negociación colectiva: resistencias vs buenas prácticas negociales 432

V. A modo de conclusiones 437

VI. Bibliografía 437

IA, EMPLEABILIDAD Y DERECHOS FUNDAMENTALES

JOSÉ VARELA FERRÍO 441

I. Presentación 441

II. Fascinación tecnológica e IA 441

III. Aplicaciones de la IA al mundo del trabajo y del empleo 444

3.1. IA y volumen de empleo 446

3.2. IA y calidad del empleo 451

IV. Conclusiones 456

V. Bibliografía 458

PARTE III.
ÉTICA, DEMOCRACIA
Y DERECHOS FUNDAMENTALES

IMPLICACIONES DE LA IA EN LA AUTONOMÍA PERSONAL Y LAS LIBERTADES DEMOCRÁTICAS

PABLO NEIRA-CASTRO Y FRANCISCO LARA 465

I. Introducción 465

II. IA y autonomía personal 466

Si quieres adquirir esta obra haz click aquí

Página

2.1. Vigilancia y privacidad 467
2.2. Toma de decisiones 469
2.3. Capacidades cognitivas y morales 475
2.4. Identidad personal 477
III. IA y democracia 479
3.1. Libertad de expresión y acceso a la información 479
3.2. Procesos democráticos y participación política 481
3.3. Igualdad y justicia 483
IV. Propuestas regulativas 484
4.1. Evaluación del impacto tecnológico 484
4.2. Fomento de la participación ciudadana y la deliberación 485
4.3. Marcos regulativos y de supervisión 487
4.3.1. Promover los valores y principios democráticos . 487
4.3.2. Establecer límites y marcos adecuados para el desarrollo y uso de nuevas tecnologías 488
4.3.3. Fomentar la colaboración y la gobernanza global 488
4.4. Dificultades por superar 489
V. Conclusión 491
VI. Bibliografía 492

REGULACIÓN Y ÉTICA PARA LA PROTECCIÓN DE LOS DERECHOS FUNDAMENTALES

IDOIA SALAZAR 501
I. Inteligencia Artificial, ventajas y oportunidades 501
II. Riesgos y retos reales de la Inteligencia Artificial 504
2.1. La Explicabilidad 505
2.2. Responsabilidad 506
2.3. El futuro del trabajo y de las prestaciones sociales 507

Si quieres adquirir esta obra haz click aquí

Página

2.4. Relaciones personas-IA 507

2.5. Concentración de poder y riqueza en unas pocas empresas muy grandes 508

2.6. El uso malicioso de la inteligencia artificial 508

III. La ética en la IA 509

3.1. La importancia del contexto cultural en la toma de decisiones 509

IV. Principios éticos de la Inteligencia Artificial 511

V. Normativa más allá de la ética 514

VI. La importancia de los sandboxes en el contexto innovación-regulación 515

6.1. Comparativa con otros sandboxes europeos y mundiales 516

VII. Bibliografía 517

PROMETEO DESENCADENADO: TORSIONES IUSFILOSÓFICAS DE LA INTELIGENCIA ARTIFICIAL

DANIEL PERES DÍAZ 519

I. Introducción 519

II. Prolegómenos metodológicos 521

III. Primera torsión: anhedonia deliberativa 522

3.1. Breve digresión sobre la inteligencia 522

3.2. Ventana de contexto y anosognosia reflexiva 523

IV. Segunda torsión: razones diabólicas 526

4.1. Sobre la opacidad epistémica 526

4.2. Fenomenología de la explicación algorítmica 530

V. Tercera torsión: anomia sistópica 533

VI. Apostilla no conclusiva 536

VII. Bibliografía 536

Si quieres adquirir esta obra haz click aquí

Página

EL IMPACTO DE LA IA EN LOS PRINCIPIOS DEMOCRÁTICOS Y LA DEMOCRACIA

MARK COECKELBERGH, OIER MENTXAKA Y FRANCISCO HERRERA 539

I. Introducción: por qué LA Inteligencia Artificial es una cuestión democrática 539

II. De herramienta tecnológica a infraestructura de poder 543

2.1. La arquitectura sociotécnica de la inteligencia artificial 543

2.2. Concentración tecnológica y poder infraestructural 544

2.3. Racionalidad algorítmica y racionalidad democrática 545

2.4. Autoridad, opacidad y legitimidad 546

III. Tres dimensiones de transformación democrática 547

IV. Riesgos y oportunidades: la naturaleza dual de la Inteligencia Artificial 550

4.1. Riesgos principales para la democracia 551

4.2. Oportunidades para el fortalecimiento democrático 554

V. Condiciones para una gobernanza democrática de la Inteligencia Artificial 557

VI. Conclusiones 559

VII. Bibliografía 560

HACIA UN MUNDO MÁS JUSTO: DESCUBRIENDO Y ABORDANDO LOS SESGOS HUMANOS Y ALGORÍTMICOS

ADRIÁN ARNAIZ-RODRÍGUEZ Y NURIA OLIVER 563

I. Introducción 563

1.1. El uso de la Inteligencia Artificial en la toma de decisiones 563

1.2 La Inteligencia Artificial fiable 566

II. Sesgos en la Inteligencia Artificial 568

2.1. Sesgos en los datos 569

2.2. Sesgos en los algoritmos 570

2.3. Sesgos en la interacción humana 572

Si quieres adquirir esta obra haz click aquí

Página

III. Métricas de sesgo y justicia algorítmica 573
3.1. Definiciones Observacionales 574
3.1.1. Métricas de justicia algorítmica de grupo 575
A. Independencia 575
B. Separación 576
C. Suficiencia 576
D. Incompatibilidades entre métricas de justicia de grupo 577
E. ¿Qué métrica de justicia algorítmica elegir en cada contexto? 579
F. Justicia vs rendimiento algorítmicos 580
3.1.2. Métricas de justicia algorítmica individual 581
3.2. Definiciones causales de justicia algorítmica 582
3.3. Definiciones de justicia en datos y aplicaciones específicas 584
IV. Técnicas de mitigación de sesgos algorítmicos 586
4.1. Mitigación Algorítmica del Sesgo 586
4.2. Interacción Humano-IA 588
4.3. Gobernanza de la Inteligencia Artificial 589
V. Conclusión 591
VI. Bibliografía 591

PARTE IV.
APLICACIONES, IMPACTO SOCIAL
Y DESARROLLOS TECNOLÓGICOS

Lenguaje, IA generativa,
propósito general y sistemas agénticos

INTELIGENCIA ARTIFICIAL Y LENGUAJE

ASUNCIÓN GÓMEZ-PÉREZ, VÍCTOR RODRÍGUEZ-DONCEL
Y PABLO CALLEJA 599

I. Introducción 599

Si quieres adquirir esta obra haz click aquí

Página

II. Inteligencia Artificial y procesamiento del lenguaje natural 602

2.1. Aproximación simbólica 603

2.2. Aproximación estadística 605

2.3. Aproximación neuronal 606

III. Marcos de evaluación 608

3.1. Evaluación de la aproximación simbólica y estadística 608

3.2. Evaluación de la aproximación neuronal 609

IV. Impacto en el lenguaje de las tecnologías del lenguaje 611

V. Impacto social de las nuevas tecnologías del lenguaje 612

5.1. Introducción 612

5.2. Tecnologías del lenguaje y sociedad 613

5.3. Cuestiones epistemológicas 616

VI. Impacto ético de las nuevas tecnologías del lenguaje 616

6.1. Sesgos y discriminación 616

6.2. Privacidad y seguridad 617

6.3. Impacto medioambiental 618

6.4. Las tecnologías bien usadas 618

VII. Impacto jurídico de las nuevas tecnologías del lenguaje 620

VIII. Conclusiones 621

INTELIGENCIA ARTIFICIAL DE PROPÓSITO GENERAL

JUAN JOSÉ HERRERA ARANDA, DANIEL MOLINA CABRERA E ISAAC TRIGUERO VELÁZQUEZ 623

I. De la IA tradicional a una IA de propósito general 623

1.1. Evolución histórica 623

1.2. Sistemas de propósito específico 624

1.3. Estado actual y perspectivas 625

II. Sistemas de IA de propósito general 625

Si quieres adquirir esta obra haz click aquí

Página

2.1. Definición de GPAIS 626

2.2 Tipos de GPAIS según sus grados de autonomía y generalidad 627

2.3 Enfoques para la construcción de GPAIS 628

2.3.1. IA potenciada por IA 628

2.4 Modelos Fundacionales (Multitarea) 630

III. Discusión sobre las implicaciones, regulación y gobernanza de GPAIS 631

3.1. Implicaciones en nuestra sociedad 631

3.2. Regulación y gobernanza en GPAIS 633

IV. Conclusiones 634

V. Bibliografía 635

DE LA GENERACIÓN A LA ACCIÓN: INTELIGENCIA ARTIFICIAL GENERATIVA Y AGÉNTICA

DAVID HERRERA-POYATOS, CARLOS PELÁEZ-GONZÁLEZ, CRISTINA ZUHEROS Y FRANCISCO HERRERA 637

I. Introducción: 2022, el año en que la Inteligencia Artificial llegó al gran público 637

II. Inteligencia Artificial generativa: de la producción de contenidos al razonamiento 641

2.1. Los Transformers y el origen de los grandes modelos de lenguaje . 641

2.2. Del modelo reutilizable a la aplicación generativa 642

2.3. Del modelo al sistema generativo 642

2.4. Fluidez, errores y alucinaciones 643

2.5. De la generación al razonamiento 644

2.6. Los nuevos modelos de razonamiento 644

2.7. Generar no es todavía actuar 645

III. Inteligencia Artificial agéntica: de los modelos que responden a los sistemas que actúan 646

Si quieres adquirir esta obra haz click aquí

Página

3.1. ¿Qué es un agente de inteligencia artificial? 646
3.2. El paradigma ReAct 647
3.3. Los componentes de un sistema agéntico 647
3.4. Agentes de IA e inteligencia artificial agéntica 649
3.5. Del agente individual a los sistemas multiagente 650
3.6. Pequeños modelos de lenguaje en los sistemas agénticos 651
3.7. Autonomía y supervisión humana 652

IV. La frontera emergente: sistemas autoevolutivos y equipos científicos artificiales 653

4.1. De los agentes adaptativos a los sistemas autoevolutivos 653
4.2. De la asistencia científica a la automatización del ciclo de investigación 655
4.3. Equipos artificiales conectados con la experimentación real 656
4.4. Autonomía epistemológica y papel del investigador humano 657
4.5. Retos del futuro agéntico 659

V. Conclusiones: de la generación a la acción 660

VI. Bibliografía 663

VISIÓN GLOBAL DE ALGUNOS RIESGOS DE LA INTELIGENCIA ARTIFICIAL GENERATIVA

MARÍA JESÚS JIMÉNEZ LINARES 667

I. Introducción 667

II. Contenido de la IAG y riesgos 670

2.1. Contenido 670
2.2. Riesgo de desinformación y confusión 670
2.3. El riesgo del ataque a los derechos al honor, la intimidad personal o familiar y la propia imagen 674
2.4. El riesgo de infracción o vulneración de los derechos de autor (propiedad intelectual) y la moderación de contenidos 675

Si quieres adquirir esta obra haz click aquí

Página

2.5. El riesgo laboral 677
2.6. El riesgo para el medio ambiente 679
2.7. El riesgo en relación con la seguridad: la protección de los datos .. 680
2.8. El riesgo por los sesgos 681
2.9. El riesgo para la ciberseguridad 684
2.10. Los riesgos psicológicos 685
2.11. Los riesgos por errores y alucinaciones 687
2.12. Los riesgos en relación a las cuestiones éticas 688
III. Clasificación de la IA generativa en función del RIA 688
IV. Conclusión 690
V. Bibliografía 692

INTELIGENCIA ARTIFICIAL SEGURA: ROBUSTEZ, PROTECCIÓN FRENTE A ATAQUES Y COMPORTAMIENTO SEGURO

ROSANA MONTES Y FRANCISCO HERRERA 697
I. Introducción 697
II. Seguridad frente a ataques y seguridad del comportamiento: dos dimensiones complementarias 701
2.1. Un concepto con dos dimensiones 701
2.2. La intención como criterio de distinción 702
2.3. Intersecciones entre ambas dimensiones 702
2.4. Una visión integrada de la seguridad 703
III. Seguridad frente a ataques y usos maliciosos 704
3.1. Prevención como primera defensa 704
3.2. Superficie de ataque y principales amenazas 704
3.3. Inyección de instrucciones y evasión de salvaguardas 706
3.4. Agentes, herramientas y control de permisos 706
3.5. Uso malicioso y red teaming 708

Si quieres adquirir esta obra haz click aquí

Página

3.6. Defensa en profundidad, respuesta y recuperación 709

IV. Robustez, evaluación y seguridad durante el ciclo de vida 710

4.1. Funcionar bien a pesar de lo desconocido 710

4.2. Robustez y principales formas de fallo 711

4.3. Evaluación mediante benchmarks y pruebas realistas 712

4.4. Evaluación mediante observabilidad y trazabilidad de agentes 713

4.5. La emergencia del paradigma Safety by Measurement 713

4.6. Observabilidad de IA: la visibilidad proactiva 715

4.7. Garantía y evaluación durante todo el ciclo de vida 716

V. Comportamiento seguro ante las personas 717

5.1. Observando la seguridad del comportamiento 717

5.2. Fiabilidad, alineamiento y control 718

5.3. Errores, persuasión y confianza excesiva 720

5.4. Agentes autónomos y supervisión humana 721

VI. Gobernanza y aseguramiento continuo de la seguridad de la Inteligencia Artificial 723

6.1. La gobernanza como elemento de la seguridad de la IA 723

6.2. El ecosistema europeo de supervisión y aseguramiento continuo 724

6.3. Perspectivas futuras: hacia una seguridad basada en observabilidad 725

VII. Conclusiones 726

VIII. Bibliografía 727

Transformación social, económica
y sostenibilidad

IA: ENEMIGA Y ALIADA EN LA GUERRA DE LA INFORMACIÓN

JAVIER MORALES BOLÍVAR 731

I. Introducción 731

II. La verdad entre el ruido 733

Si quieres adquirir esta obra haz click aquí

Página

III. La IA como aliada 737

3.1. La regulación 737

3.2. Mejorar el trabajo y fortalecer la credibilidad 741

3.3. Al revés: el periodismo al servicio de la IA 745

IV. Conclusiones 746

V. Bibliografía 747

UN ROBOT EN MI OFICINA

SENÉN BARRO AMENEIRO 753

I. Hágase la máquina y haga nuestro trabajo 753

1.1. Primeros pasos: de los animales a las máquinas 753

1.2. Revoluciones industriales 754

1.3. De la automatización física a la cognitiva 755

II. Y llegó la automatización inteligente 755

2.1. La automatización basada en IA y su repercusión en el trabajo humano 755

2.2. Cambios en los perfiles profesionales 758

III. Crecimiento económico y aumento de la productividad 763

3.1. ¡Es la innomatización, estúpido! 765

IV. El día de mañana en un mundo de personas y máquinas 767

4.1. No hay determinismo en el uso de la tecnología 768

V. A modo de cierre 769

GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL: PROFUNDIZANDO EN LOS RIESGOS

REYES CALDERÓN Y JAVIER FUERTES 771

I. Introducción 771

II. Riesgos de la IA 773

Si quieres adquirir esta obra haz click aquí

Página

2.1. Esquema clásico de riesgos tecnológicos 775

2.2. Riesgos estructurales 776

2.3. Riesgos estructurales procedentes de la IA per se 777

III. Referencias 779

INTELIGENCIA ARTIFICIAL VERDE: ENFOQUES, REGULACIÓN Y PERSPECTIVAS PARA LA SOSTENIBILIDAD

AMPARO ALONSO BETANZOS, VERÓNICA BOLÓN CANEDO, LAURA MORÁN FERNÁNDEZ Y BRAIS CANCELA BARIZO 785

I. Introducción 785

II. Green-by IA 790

2.1. Eficiencia energética 790

2.2. Movilidad inteligente 791

2.3. Agricultura sostenible 792

2.4. Cambio climático 794

2.5. Políticas medioambientales 795

III. Green-in IA 795

3.1. Optimización algorítmica 796

3.2. Optimización del hardware 797

3.3. El rol de los centros de datos y la energía 797

3.4. Reducción pragmática del factor de escala 798

IV. Herramientas para el cálculo del consumo energético 799

V. Regulación e Inteligencia Artificial verde 800

VI. Orientaciones futuras 805

SOBRE EL FUTURO DE LA INTELIGENCIA ARTIFICIAL Y SU POTENCIAL TRANSFORMADOR PARA NUESTRA CIVILIZACIÓN

JESÚS L. LOBO Y JAVIER DEL SER 809

I. Introducción 809

Si quieres adquirir esta obra haz click aquí

Página

II. La necesidad de una perspectiva ética de la IA 811

2.1. Perspectivas éticas de la IA 812

2.1.1. Utilitarismo 812

2.1.2. Deontología 813

2.1.3. Ética de las virtudes 813

2.1.4. Contractualismo 813

2.1.5. Ética del cuidado 814

2.2. Desafíos en la aplicación de la ética a la IA 814

III. La IA hasta ahora 815

IV. De la IA actual a la IA transformadora 817

4.1. Factores humanos 818

4.1.1. La sobrecarga de información 818

4.1.2. El desalineamiento 819

4.1.3. La percepción negativa y la desconfianza respecto a la IA 819

4.1.3.1. La "economía de la atención" 820

4.1.3.2. La "economía de la intención" 820

4.1.3.3. Humano vs IA 821

4.1.4. Excesiva dependencia de la experiencia humana 821

4.1.5. La percepción de futilidad de la IA ética 821

4.1.6. La pérdida de confianza y la resistencia 822

4.1.7. Una IA "diluida" 823

4.1.8. La sobrerregulación 824

4.1.9. La obsolescencia del ser humano 824

4.1.10. La "IA-cracia" 825

4.2. Factores científico-tecnológicos 826

4.2.1. La paradoja de los datos 826

4.2.2. La dificultad de detectar la aparición de nuevas capacidades 827

Si quieres adquirir esta obra haz click aquí

Página

4.2.3. Modelando el mundo 828

4.2.4. Sostenibilidad, limitaciones físicas y alternativas 828

4.2.5. La IA, el nicho de unos pocos 829

4.2.6. La dualidad de los fundamentos teóricos de la computación y la IA..... 830

4.2.6.1. La mejora recursiva: entre la promesa y el riesgo sistémico..... 831

4.2.7. La seguridad de la IA 832

V. Oportunidades para una IA transformadora..... 833

5.1. Algunos brotes verdes en el camino hacia la IA transformadora... 833

5.2. La Gran Esperanza: una "explosión" científica..... 837

VI. ¿Necesitaremos una nueva perspectiva ética y filosófica para esta IA transformadora?..... 838

VII. Conclusiones..... 839

VIII. Bibliografía 839

Si quieres adquirir esta obra haz click aquí

de las instituciones responsables de su utilización. En cambio, la confiabilidad se refiere a las propiedades, garantías y mecanismos que proporcionan razones para depositar esa confianza. El objetivo no debe ser maximizar indiscriminadamente la aceptación de la IA, sino desarrollar sistemas, organizaciones e instituciones merecedores de una confianza crítica y justificada.

La perspectiva sociotécnica refuerza este planteamiento. Los resultados y efectos de la IA no dependen exclusivamente del código o de los datos, sino de su interacción con personas, organizaciones, mercados e instituciones. La discriminación, la transformación del empleo, la desigualdad, la pérdida de autonomía profesional y la concentración de poder tecnológico muestran que la confiabilidad es una propiedad del sistema sociotécnico completo y de su gobernanza.

A partir de estas premisas, la pregunta que guía el capítulo es la siguiente: ¿cómo pueden integrarse la reflexión ética, las garantías jurídicas y el análisis sociotécnico para desarrollar sistemas de IA confiables y generar una confianza social justificada?

Sobre estas premisas, el capítulo se estructura en tres ámbitos complementarios. En primer lugar, se examinan los fundamentos éticos de la IA y la necesidad de traducir los principios generales en prácticas, responsabilidades y mecanismos efectivos de rendición de cuentas. En segundo lugar, se analiza la respuesta jurídica europea, con especial atención al AI Act, al régimen aplicable a la IA de propósito general y a su evolución mediante el Digital Omnibus on AI. Finalmente, se estudia la IA como un fenómeno sociotécnico y socioeconómico, atendiendo a sus efectos sobre el empleo, la discriminación, la autonomía profesional, la desigualdad y la concentración de poder. A través de esta aproximación multidisciplinar, se sostiene que una confianza justificada en la IA solo puede construirse mediante la articulación de valores éticos, garantías jurídicas, requisitos técnicos e instituciones capaces de supervisar sus consecuencias en todos los ámbitos sobre los que la IA despliega sus efectos.

II. ÉTICA E INTELIGENCIA ARTIFICIAL: DE LOS PRINCIPIOS A LA RESPONSABILIDAD

La ética de la IA analiza cómo el diseño, el desarrollo y el uso de sistemas algorítmicos afectan a la autonomía, la dignidad, la justicia, la privacidad y el bienestar de las personas. Su objeto no consiste únicamente en determinar qué puede hacer técnicamente una IA, sino también en examinar qué usos resultan legítimos, qué decisiones no deberían delegarse, qué valores deben orientar su desarrollo y quién debe responder por sus consecuencias. Para entender esta posición baste con recordar los límites éticos que establecieron en torno a la clonación genética.

Si quieres adquirir esta obra haz click aquí

Coeckelbergh (2020) ofrece una introducción especialmente valiosa a esta problemática al situar la ética de la IA dentro de la relación más amplia entre tecnología, seres humanos y sociedad. Su planteamiento permite superar una visión reducida de la ética centrada exclusivamente en escenarios futuros, como la posible aparición de máquinas plenamente autónomas o moralmente responsables. Aunque estas cuestiones poseen cierto interés filosófico, la mayoría de los problemas éticos de la IA son ya actuales y se manifiestan en ámbitos como la vigilancia, la privacidad, la discriminación, la automatización del trabajo, la manipulación informativa, la toma algorítmica de decisiones y la concentración de poder tecnológico.

Desde esta perspectiva, la IA no puede considerarse una herramienta completamente neutral. Los sistemas de IA son diseñados para alcanzar determinados objetivos, utilizan categorías y datos previamente seleccionados y nacen de instituciones privadas que poseen intereses, normas e incentivos específicos. Las decisiones acerca de qué problema debe automatizarse, qué variables deben incluirse, qué errores se consideran aceptables o qué grupos soportan los riesgos tienen una dimensión ética incluso antes de que el sistema sea desplegado.

Coeckelbergh subraya asimismo que las tecnologías no solo producen resultados, sino que transforman las relaciones humanas y las formas de interpretar el mundo. La incorporación de la IA puede modificar la relación entre profesionales y usuarios, alterar la autonomía de las personas, redistribuir autoridad y crear nuevas dependencias respecto de organizaciones e infraestructuras tecnológicas. La pregunta ética no se limita, por tanto, a determinar si una decisión algorítmica es correcta, sino qué debe examinar, qué tipo de prácticas, relaciones e instituciones contribuye a configurar.

Este enfoque también ayuda a cuestionar una separación rígida entre la tecnología y la sociedad. Los valores no se incorporan únicamente al final del proceso de desarrollo, mediante un código ético o una evaluación posterior. Están presentes desde la definición de la finalidad del sistema, la elección de los datos y los criterios de optimización hasta las condiciones de despliegue, supervisión y retirada. La ética debe acompañar, en consecuencia, todo el ciclo de vida de la IA.

La literatura internacional muestra una convergencia alrededor de principios como la transparencia, la justicia, la prevención del daño, la privacidad, la autonomía y la responsabilidad. El análisis de Jobin, Ienca y Vayena (2019) evidencia que numerosos marcos éticos coinciden en estos valores generales. Sin embargo, esa convergencia terminológica no implica necesariamente una interpretación común ni una aplicación homogénea. Los principios pueden adquirir significados diferentes según el sector, el contexto institucional o las personas afectadas.

Si quieres adquirir esta obra haz click aquí

En el ámbito europeo, las directrices para una IA confiable desarrollaron estas preocupaciones mediante siete requisitos: agencia y supervisión humana; robustez técnica y seguridad; privacidad y gobernanza de datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental; y rendición de cuentas (High-Level Expert Group on Artificial Intelligence, 2019). Estos requisitos muestran que la confiabilidad no se limita a la precisión del modelo, sino que comprende también la calidad de los procesos, la protección de las personas y la responsabilidad de las organizaciones.

No obstante, la formulación de principios no garantiza por sí sola un comportamiento ético. Mittelstadt (2019) señala que conceptos como justicia, autonomía o transparencia pueden ser interpretados de formas distintas e incluso entrar en conflicto. Por ejemplo, aumentar la precisión de un sistema puede requerir una utilización más intensiva de datos personales; una transparencia completa puede afectar a la seguridad o a la protección de información legítimamente reservada; y distintas concepciones de la equidad pueden producir resultados incompatibles entre sí.

Los principios generales tampoco indican automáticamente el modo de resolver estos conflictos, quién tiene autoridad para adoptar las decisiones o qué consecuencias deben derivarse de su incumplimiento. Esta indeterminación puede favorecer el denominado *ethics washing*, mediante el cual una organización adopta declaraciones éticas o principios voluntarios sin modificar sustancialmente sus procesos de desarrollo, sus incentivos o sus mecanismos de responsabilidad.

El principal desafío consiste, por tanto, en traducir los principios aceptados generalmente en prácticas concretas observables y verificables. Esto requiere asignar responsabilidades, documentar las decisiones, realizar evaluaciones de impacto, incorporar mecanismos de supervisión humana y permitir la participación de las personas potencialmente afectadas. También exige establecer procedimientos para detectar daños, revisar resultados, gestionar incidentes y proporcionar vías de reclamación y reparación.

La revisión de alcance realizada por Sadek et al. (2025) profundiza precisamente en la distancia existente entre los compromisos generales de IA responsable y su aplicación en contextos reales. Los autores muestran que la dificultad no reside únicamente en la ausencia de principios éticos, sino en la falta de condiciones organizativas que permitan trasladarlos a la práctica. Entre los principales obstáculos se encuentran el carácter abstracto y no vinculante de muchas directrices, la inexistencia de métricas consensuadas, los conflictos entre valores, la fragmentación del proceso de desarrollo y la distribución imprecisa de responsabilidades.

Si quieres adquirir esta obra haz click aquí

Un problema relevante es que las diferentes fases del ciclo de vida suelen estar separadas entre equipos especializados. Quienes definen los objetivos del sistema, recopilan los datos, desarrollan el modelo, realizan la validación o deciden su despliegue pueden pertenecer a departamentos distintos y manejar criterios de éxito diferentes. Esta fragmentación dificulta identificar quién debe asumir la responsabilidad por los efectos finales y puede provocar que los problemas éticos sean considerados competencia de otros actores.

Sadek et al. también ponen de manifiesto la tensión entre los objetivos de responsabilidad y otras prioridades organizativas, como la rapidez de lanzamiento, la reducción de costes, el rendimiento comercial o la mejora de las prestaciones técnicas. En ausencia de estructuras formales de gobernanza, los principios éticos pueden quedar subordinados a incentivos más inmediatos. La implantación de una IA responsable requiere, por ello, apoyo directivo, recursos suficientes, formación, coordinación interdisciplinar y procedimientos que integren las consideraciones éticas en las decisiones ordinarias de la organización.

La revisión propone avanzar hacia actuaciones más contextualizadas y operativas. No existe una única técnica capaz de garantizar la responsabilidad de todos los sistemas de IA. Las medidas deben adaptarse al sector, la finalidad, el nivel de riesgo y las necesidades de las partes interesadas. Esto implica combinar herramientas técnicas —como pruebas de robustez, análisis de sesgos o métodos de explicabilidad— con medidas organizativas, entre ellas evaluaciones de impacto, protocolos de documentación, comités de supervisión, mecanismos de participación y sistemas de auditoría.

La contribución de Sadek et al. resulta especialmente importante porque confirma que los problemas éticos de la IA poseen una naturaleza sociotécnica. Un resultado injusto puede estar relacionado con los datos o el modelo, pero también con una finalidad mal definida, una utilización fuera del contexto previsto, una interpretación incorrecta de las predicciones o una supervisión humana meramente formal. Las soluciones deben considerar, por tanto, la interacción entre componentes técnicos, decisiones humanas y estructuras institucionales.

En consecuencia, la ética de la IA no debe entenderse como un catálogo estático de valores ni como una fase añadida al final del desarrollo. Constituye una práctica continuada de deliberación, justificación y responsabilidad. Los principios éticos ofrecen la orientación inicial, pero solo contribuyen a una IA confiable cuando se traducen en requisitos concretos, decisiones documentadas, prácticas organizativas y mecanismos efectivos de rendición de cuentas.

Esta transición desde los principios hacia la responsabilidad institucional permite, además, conectar la ética con las restantes dimensiones del capítulo. Algunas exigencias éticas han sido transformadas en obligaciones jurídicas mediante el AI

Si quieres adquirir esta obra haz click aquí

Act, mientras que otras requieren un análisis más amplio de los contextos organizativos y sociales en los que se utiliza la IA. La ética proporciona así el fundamento normativo desde el cual evaluar tanto la regulación como la gobernanza sociotécnica de estos sistemas.

III. DERECHO E INTELIGENCIA ARTIFICIAL: AI ACT Y SU EVOLUCIÓN MEDIANTE EL DIGITAL OMNIBUS ON AI

El desarrollo acelerado de la IA ha evidenciado las limitaciones de los instrumentos exclusivamente éticos o voluntarios. Los principios permiten orientar el diseño y la utilización de los sistemas, pero no establecen por sí mismos obligaciones exigibles, mecanismos de supervisión ni consecuencias jurídicas ante el incumplimiento. El Derecho desempeña, por ello, una función esencial al transformar determinadas exigencias de seguridad, transparencia, equidad y responsabilidad en deberes jurídicamente vinculantes.

El Reglamento 2024/1689, conocido como Reglamento de Inteligencia Artificial o AI Act, constituye el principal marco normativo de la UE en esta materia. Su objetivo es armonizar las condiciones para la introducción en el mercado, la puesta en servicio y la utilización de sistemas de IA, promoviendo la innovación y garantizando un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales.

El AI Act debe entenderse, no obstante, como un marco en evolución. En noviembre de 2025, la Comisión Europea presentó el Digital Omnibus on AI (Digital Omnibus, 2025), una propuesta orientada a simplificar y facilitar la aplicación de algunas de sus disposiciones. Tras el acuerdo interinstitucional, el Parlamento Europeo aprobó el texto el 16 de junio de 2026, aunque en el momento de redactar este capítulo todavía estaba pendiente de adopción formal por el Consejo. Por tanto, debe distinguirse entre las obligaciones ya vigentes del AI Act y las modificaciones cuya entrada en vigor depende de la culminación del procedimiento legislativo.

3.1. UNA ARQUITECTURA JURÍDICA BASADA EN LOS RIESGOS

El AI Act establece un régimen diferenciado según la finalidad, el contexto y el nivel de riesgo de cada sistema. Su estructura distingue principalmente entre prácticas prohibidas, sistemas de alto riesgo, sistemas sometidos a obligaciones de transparencia y modelos de IA de propósito general.

Las prácticas prohibidas corresponden a usos considerados incompatibles con los valores y derechos protegidos por la Unión Europea. Su inclusión demuestra que la regulación no se limita a gestionar técnicamente todos los riesgos. Algunas

Si quieres adquirir esta obra haz click aquí

aplicaciones se consideran jurídicamente inadmisibles, con independencia de que su precisión o seguridad puedan mejorarse.

Los sistemas de alto riesgo constituyen el núcleo del Reglamento. Esta categoría comprende determinadas aplicaciones utilizadas en ámbitos como la biometría, las infraestructuras críticas, la educación, el empleo, el acceso a servicios esenciales, la aplicación de la ley, la migración o la administración de justicia. También incluye sistemas que funcionan como componentes de seguridad de productos regulados por la legislación europea.

La clasificación no depende exclusivamente del modelo empleado. Un mismo componente tecnológico puede quedar sujeto a regímenes diferentes según la finalidad para la que se utilice y las consecuencias que pueda producir. El riesgo jurídico emerge, por tanto, de la relación entre las capacidades técnicas, el contexto institucional y las personas eventualmente afectadas.

Los proveedores de sistemas de alto riesgo deben cumplir requisitos relacionados con la gestión continua de riesgos, la gobernanza de datos, la documentación técnica, el registro de eventos, la transparencia, la supervisión humana, la precisión, la robustez y la ciberseguridad. Estas obligaciones se aplican durante todo el ciclo de vida y no únicamente en el momento de introducir el sistema en el mercado.

La supervisión humana resulta especialmente relevante, puesto que el sistema debe permitir que una persona comprenda sus capacidades y limitaciones, identifique anomalías y disponga de autoridad para ignorar, modificar o detener sus resultados. Sin embargo, la mera presencia formal de una persona no garantiza una intervención efectiva. Son necesarias condiciones organizativas que proporcionen conocimientos, tiempo, información y capacidad real de decisión.

El régimen de alto riesgo convierte así varias propiedades asociadas a la IA confiable en obligaciones jurídicas. La confiabilidad deja de depender exclusivamente de compromisos voluntarios y pasa a requerir evidencias documentadas, procedimientos de evaluación y mecanismos de seguimiento.

3.2. RESPONSABILIDADES, TRANSPARENCIA Y PROTECCIÓN DE LAS PERSONAS AFECTADAS

El AI Act distribuye las obligaciones entre los distintos actores de la cadena de valor, entre ellos proveedores, importadores, distribuidores y responsables del despliegue. Esta distribución responde a la constatación de que los riesgos pueden originarse en diferentes fases del diseño, comercialización y utilización del sistema.

El proveedor controla normalmente el desarrollo, el entrenamiento, la validación y la documentación. El responsable del despliegue decide cómo integrar el sistema

Si quieres adquirir esta obra haz click aquí

en una organización, sobre qué personas aplicarlo y cómo interpretar sus resultados. Ambos poseen capacidades distintas para prevenir, identificar y corregir los daños.

Los proveedores deben establecer sistemas de gestión de calidad, realizar las evaluaciones de conformidad correspondientes, conservar la documentación y desarrollar mecanismos de seguimiento posterior a la comercialización. Los responsables del despliegue deben utilizar el sistema conforme a las instrucciones, garantizar una supervisión humana adecuada, controlar los datos de entrada bajo su responsabilidad y comunicar los incidentes relevantes.

En determinados supuestos, los responsables del despliegue de sistemas de alto riesgo deben realizar una evaluación de impacto sobre los derechos fundamentales. Esta evaluación permite considerar las características concretas del contexto, las personas y grupos afectados, los riesgos previsibles y las medidas de supervisión, reclamación y reparación.

El Reglamento establece también obligaciones de transparencia para ciertos sistemas que no se clasifican necesariamente de alto riesgo. Las personas deben ser informadas cuando interactúan con determinados sistemas de IA y deben poder reconocer ciertos contenidos generados o manipulados artificialmente.

La transparencia no puede reducirse, sin embargo, a una comunicación formal. La información debe ser accesible, comprensible y proporcionada a las necesidades de cada destinatario. Una explicación excesivamente técnica puede cumplir documentalmente una obligación sin permitir que la persona comprenda las consecuencias de la utilización del sistema.

La información adquiere valor jurídico cuando se conecta con posibilidades efectivas de revisión e impugnación. La protección de las personas afectadas exige no solo conocer que se utiliza IA, sino también disponer de mecanismos para cuestionar decisiones, solicitar explicaciones pertinentes y obtener reparación cuando se produce un daño.

3.3. LA REGULACIÓN DE LOS MODELOS DE INTELIGENCIA ARTIFICIAL DE PROPÓSITO GENERAL

Los modelos de IA de propósito general plantean un desafío específico porque no se desarrollan necesariamente para una aplicación concreta. Pueden utilizarse directamente o integrarse posteriormente en numerosos sistemas, productos y servicios.

Esta versatilidad dificulta una regulación basada exclusivamente en el uso final. El proveedor del modelo puede desconocer todos los contextos en los que este será utilizado, mientras que quienes lo integran en aplicaciones posteriores pueden

Si quieres adquirir esta obra haz click aquí

carecer de información suficiente sobre sus capacidades, limitaciones y condiciones de entrenamiento.

El AI Act establece obligaciones para los proveedores de modelos de propósito general. Estas incluyen la elaboración de documentación técnica, la transmisión de información a los proveedores posteriores, la adopción de una política para cumplir la legislación europea sobre derechos de autor y la publicación de un resumen suficientemente detallado del contenido utilizado durante el entrenamiento.

El Reglamento diferencia además los modelos que pueden presentar riesgos sistémicos. Sus proveedores deben evaluar y mitigar esos riesgos, documentar y comunicar incidentes graves y garantizar un nivel adecuado de ciberseguridad. El concepto de riesgo sistémico desplaza la atención desde una aplicación aislada hacia los efectos que un modelo avanzado puede producir en múltiples sectores, sistemas e infraestructuras.

La aplicación práctica de este régimen se apoya en las directrices de la Comisión sobre las obligaciones de los proveedores y en el Código de Buenas Prácticas para la IA de propósito general. Este último constituye un instrumento voluntario destinado a facilitar el cumplimiento en materia de transparencia, derechos de autor y seguridad. Las obligaciones legales aplicables a los nuevos modelos de propósito general comenzaron a aplicarse el 2 de agosto de 2025.

Debe distinguirse entre el modelo de propósito general y el sistema desarrollado a partir de él. El modelo proporciona capacidades generales, mientras que el sistema incorpora esas capacidades en una aplicación, interfaz o proceso organizativo concreto. Esta distinción es esencial para determinar qué información debe circular a lo largo de la cadena de valor y qué responsabilidades corresponden al proveedor del modelo, al desarrollador posterior y al responsable del despliegue.

Como señalan Gstrein et al. (2024), la regulación de estos modelos representa un avance hacia una gobernanza preventiva, pero plantea dificultades relacionadas con la delimitación de las obligaciones, la evolución rápida de las capacidades técnicas y la distribución de responsabilidades entre múltiples actores.

3.4. EL DIGITAL OMNIBUS Y LA GOBERNANZA INSTITUCIONAL DEL AI ACT

El Digital Omnibus on AI persigue simplificar la aplicación de determinadas disposiciones del AI Act, reducir solapamientos regulatorios y proporcionar mayor seguridad jurídica. La reforma no sustituye la arquitectura basada en riesgos ni elimina el régimen aplicable a los modelos de propósito general, sino que introdu-

Si quieres adquirir esta obra haz click aquí

ce ajustes en los plazos, los procedimientos de cumplimiento y la distribución de competencias.

El texto aprobado por el Parlamento Europeo prevé aplazar la aplicación de determinados requisitos de los sistemas de alto riesgo. Las obligaciones relativas a sistemas de alto riesgo independientes se aplicarían a partir del 2 de diciembre de 2027, mientras que las correspondientes a sistemas integrados como componentes de seguridad de productos sujetos a legislación sectorial lo harían desde el 2 de agosto de 2028. También se prevé retrasar hasta el 2 de diciembre de 2026 algunas obligaciones de marcado de contenidos generados mediante IA.

Estos cambios responden, principalmente, a la necesidad de disponer de estándares armonizados, orientaciones y herramientas técnicas suficientes para demostrar el cumplimiento. El aplazamiento puede favorecer la seguridad jurídica, pero plantea también el riesgo de retrasar salvaguardias relevantes en sectores sensibles. La simplificación debe evaluarse, por tanto, atendiendo a su capacidad para facilitar el cumplimiento sin reducir el nivel de protección.

La reforma adquiere especial importancia para los sistemas construidos sobre modelos de propósito general. El texto aprobado pretende centralizar en mayor medida la supervisión de determinados sistemas cuando el modelo general y la aplicación derivada han sido desarrollados por el mismo proveedor. Esta función correspondería principalmente a la Oficina Europea de Inteligencia Artificial, sin perjuicio de ciertas competencias nacionales y sectoriales.

La gobernanza institucional del AI Act se distribuye entre la Oficina Europea de IA, el Consejo Europeo de Inteligencia Artificial, el panel científico y las autoridades nacionales. Novelli et al. (2025) subrayan que la eficacia de esta estructura dependerá de una distribución clara de competencias, la independencia de los órganos supervisores, la disponibilidad de conocimientos especializados y la coordinación entre los niveles europeo y nacional.

La centralización puede reducir la fragmentación y facilitar una supervisión coherente de modelos utilizados en varios Estados miembros. Al mismo tiempo, incrementa las responsabilidades de la Oficina Europea de IA y exige que disponga de recursos suficientes para evaluar sistemas desarrollados por empresas con una capacidad técnica y económica considerable.

El Digital Omnibus confirma así que la efectividad del AI Act no depende únicamente de la calidad de sus disposiciones. También requiere estándares aplicables, autoridades competentes, mecanismos de cooperación y capacidad pública para verificar el cumplimiento.

Si quieres adquirir esta obra haz click aquí

3.5. DEL CUMPLIMIENTO JURÍDICO A LA INTELIGENCIA ARTIFICIAL CONFIABLE

El AI Act constituye un avance decisivo porque convierte determinadas exigencias de la IA confiable en obligaciones vinculantes. La gestión de riesgos, la gobernanza de datos, la documentación, la transparencia, la supervisión humana y la robustez dejan de ser únicamente recomendaciones y pasan a integrarse en un régimen de evaluación, vigilancia y responsabilidad.

Sin embargo, el cumplimiento jurídico no equivale automáticamente a legitimidad ética o social. Una aplicación puede cumplir formalmente los requisitos aplicables y continuar planteando dudas sobre su necesidad, proporcionalidad, impacto laboral o distribución de beneficios y riesgos. El Derecho establece garantías mínimas, pero no en pocas ocasiones puede anticipar los conflictos o sustituir la deliberación ética.

Tampoco debe reducirse el cumplimiento a una lista documental. La gestión de riesgos debe reflejar las condiciones reales de utilización; la supervisión humana debe ser efectiva; y la documentación debe corresponder con las prácticas desarrolladas por la organización. La conformidad formal pierde valor cuando no existe una correspondencia con el funcionamiento real del sistema.

El AI Act debe aplicarse, además, junto con otras normas relativas a protección de datos, consumo, seguridad de productos, no discriminación, trabajo, competencia, propiedad intelectual y responsabilidad. Los problemas derivados de la IA no se agotan en una única norma, especialmente cuando afectan a estructuras económicas, relaciones laborales o daños sociales acumulativos.

La relación entre ética y Derecho debe entenderse, por tanto, como complementaria. La ética permite examinar qué finalidades deben perseguirse y qué responsabilidades existen más allá del mínimo legal. El Derecho proporciona obligaciones exigibles, instituciones supervisoras y mecanismos de protección. La perspectiva sociotécnica permite comprobar cómo esas normas y valores operan dentro de organizaciones y contextos concretos.

En consecuencia, el AI Act es una condición esencial, pero no suficiente, para desarrollar una IA confiable. Su efectividad dependerá de la aplicación coherente de sus obligaciones, de la capacidad de sus instituciones y de su integración con una evaluación ética y sociotécnica más amplia. La confianza justificada no surge únicamente del cumplimiento formal, sino de la existencia de una cadena verificable de garantías, responsabilidades y mecanismos de supervisión durante todo el ciclo de vida del sistema.

Si quieres adquirir esta obra haz click aquí

IV. INTELIGENCIA ARTIFICIAL Y SOCIEDAD

La IA no puede comprenderse adecuadamente como un conjunto aislado de modelos, algoritmos y datos. Los sistemas de IA se desarrollan y utilizan dentro de organizaciones, mercados, instituciones públicas y relaciones sociales que determinan sus objetivos, condiciones de funcionamiento y efectos. Su impacto depende tanto de sus propiedades técnicas como de las decisiones humanas, los incentivos económicos, las estructuras de poder y los contextos institucionales en los que se integran.

Esta perspectiva sociotécnica permite ampliar la unidad de análisis. El objeto que debe evaluarse no es únicamente el modelo computacional, sino el sistema completo compuesto por datos, infraestructuras, desarrolladores, empresas, trabajadores, usuarios, personas afectadas, autoridades públicas y normas sociales y jurídicas. Un mismo modelo puede producir consecuencias muy diferentes según quién lo utilice, con qué finalidad, sobre qué población y bajo qué mecanismos de supervisión.

La perspectiva sociotécnica debe complementarse, además, con una dimensión socioeconómica. La IA transforma la productividad, el empleo, la distribución de la renta, la organización empresarial y la estructura de los mercados. También puede modificar el equilibrio entre capital y trabajo y concentrar capacidades tecnológicas y económicas en un número reducido de organizaciones. La cuestión fundamental no es solamente cuánto progreso técnico puede generar la IA, sino cómo se orienta ese progreso y quién recibe sus beneficios.

4.1. LA INTELIGENCIA ARTIFICIAL COMO SISTEMA SOCIOTÉCNICO

Johnson y Verdicchio[6] distinguen entre la consideración de la IA como artefacto computacional y su comprensión como sistema sociotécnico. En la primera perspectiva, el sistema se identifica principalmente con su código, sus modelos y sus datos. Desde la segunda, la IA está formada por componentes técnicos y sociales que se encuentran profundamente interrelacionados.

Esta distinción modifica la forma de analizar los valores asociados a la tecnología. Cuando la IA se presenta como un artefacto neutral, puede parecer que principios como la justicia, la autonomía o la transparencia pueden añadirse posteriormente al sistema mediante requisitos técnicos. Sin embargo, estos valores no constituyen propiedades unívocas que puedan traducirse automáticamente en

6. JOHNSON, D. G. y VERDICCHIO, M., *The Sociotechnical Entanglement of AI and Values*, *AI & Society*, vol. 40, 2025, pp. 67-76.

Si quieres adquirir esta obra haz click aquí

código. Su interpretación depende del contexto, de las instituciones y de las personas afectadas.

La justicia, por ejemplo, puede referirse a la igualdad de trato, la igualdad de oportunidades, la compensación de desigualdades históricas o la distribución equitativa de beneficios y riesgos. La selección de una interpretación no es una decisión exclusivamente matemática. Responde *a priori*dades normativas y sociales que deben ser justificadas.

Johnson y Verdicchio sostienen que la relación entre la IA y los valores debe comprenderse como un proceso de coproducción. Las instituciones y las prácticas sociales influyen en la tecnología, pero la tecnología también transforma esas instituciones y prácticas. Los valores no se encuentran únicamente dentro del algoritmo, sino que se manifiestan en la definición del problema, la selección de datos, los criterios de evaluación, las condiciones de uso y la distribución de responsabilidades.

Una organización que desarrolla un sistema de selección de personal, por ejemplo, debe decidir qué entiende por un candidato adecuado, qué características considera relevantes y qué errores está dispuesta a aceptar. Estas decisiones expresan una determinada concepción de la eficiencia, el mérito y la igualdad de oportunidades. El modelo materializa posteriormente esas prioridades, pero no las crea de forma autónoma.

La perspectiva sociotécnica permite también reconocer la diversidad de actores afectados. Los efectos de una aplicación no recaen solo sobre quienes interactúan directamente con ella. Pueden afectar a personas clasificadas por el sistema, a grupos excluidos de determinadas oportunidades, a trabajadores cuya actividad es reorganizada y a comunidades sobre las que se distribuyen sus costes sociales y económicos.

Riesen[7] profundiza en esta aproximación al cuestionar la concepción de la IA como agente autónomo que sustituye a los seres humanos. Los sistemas actuales deben entenderse como configuraciones sociotecnológicas en las que participan modelos, personas, organizaciones e infraestructuras. Incluso las aplicaciones con un elevado grado de automatización dependen de decisiones humanas sobre su finalidad, entrenamiento, validación y utilización.

Esta interdependencia permite hablar de una agencia distribuida. Los resultados de la IA emergen de la interacción entre diversos componentes y actores, y no de la actuación aislada de una máquina. La expresión "la IA tomó la decisión" puede ocultar la cadena de elecciones humanas y organizativas que configuró el sistema.

7. JOBIN, A., IENCA, M. y VAYENA, E., "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, vol. 1, 2019, pp. 389-399.

Si quieres adquirir esta obra haz click aquí

El análisis de la responsabilidad debe reconstruir esa cadena: quién decidió automatizar la tarea, quién definió los objetivos, quién seleccionó los datos, quién validó el modelo y quién tiene autoridad para intervenir cuando aparecen consecuencias perjudiciales. La distribución de la agencia no elimina la responsabilidad, sino que exige identificarla y organizarla de manera más precisa.

4.2. DIRECCIÓN DEL CAMBIO TECNOLÓGICO, PODER Y PROGRESO

La dimensión socioeconómica de la IA puede abordarse a partir de la tesis central de Acemoglu y Johnson[8]: el progreso tecnológico no conduce de forma automática a una prosperidad compartida. Los beneficios de una innovación dependen de la dirección que adopte el cambio técnico, de las instituciones que lo gobiernen y del poder de los distintos grupos para influir en su desarrollo.

En *Power and Progress*, los autores cuestionan una concepción determinista según la cual toda innovación que aumenta la productividad termina necesariamente mejorando el bienestar general. La historia muestra que el crecimiento de la capacidad productiva puede coexistir durante largos periodos con salarios estancados, peores condiciones laborales o una concentración creciente de la riqueza.

La tecnología abre posibilidades, pero las instituciones determinan cómo se distribuyen sus beneficios. Cuando las decisiones sobre innovación se concentran en un reducido grupo de empresas y actores económicos, la tecnología puede orientarse prioritariamente hacia la reducción de costes, la sustitución de trabajadores, la vigilancia o la captación de rentas. Una trayectoria alternativa puede dirigirla hacia la creación de nuevas tareas, el fortalecimiento de las capacidades humanas y la mejora de servicios histórica y socialmente valiosos.

Esta distinción resulta central para comprender el impacto de la IA. El debate no debe plantearse simplemente como una oposición entre aceptar o rechazar la automatización. La cuestión es qué tipo de automatización se promueve y qué alternativas tecnológicas quedan desplazadas. Una aplicación puede sustituir tareas humanas sin mejorar significativamente la calidad del producto, mientras que otra puede proporcionar información y herramientas que incrementen la capacidad de los trabajadores.

Acemoglu y Johnson denominan *machine usefulness* a una orientación de la innovación que aumenta las capacidades y contribuciones de las personas. Desde esta perspectiva, la IA debería emplearse para proporcionar conocimientos, apo-

8. ACEMOGLU, D. y JOHNSON, S., *Power and Progress. Our Thousand-Year Struggle over Technology and Prosperity*, Nueva York, PublicAffairs, 2023.

Si quieres adquirir esta obra haz click aquí

yar decisiones y crear nuevas tareas, en lugar de limitarse a reproducir mediante máquinas actividades que anteriormente realizaban trabajadores.

La dirección de la innovación no es una consecuencia inevitable de la tecnología. Depende de políticas públicas, incentivos fiscales, normas laborales, regulación de la competencia, inversión educativa y capacidad de participación social. Por ello, la gobernanza de la IA debe considerar no solo los riesgos de cada sistema, sino también qué trayectorias de innovación reciben apoyo y qué intereses determinan su desarrollo.

4.3. PRODUCTIVIDAD, EMPLEO Y DISTRIBUCIÓN DE LA RENTA

En un trabajo reciente de Acemoglu[9] se introduce una evaluación más prudente de los efectos macroeconómicos de la IA. Frente a las predicciones que anticipan incrementos extraordinarios e inmediatos de la productividad, el autor propone estimar esos efectos considerando qué proporción de las tareas puede ser afectada y qué reducciones de costes o mejoras productivas pueden alcanzarse en cada una.

Su análisis diferencia entre tareas relativamente sencillas, en las que el rendimiento puede observarse y aprenderse a partir de resultados claros y tareas complejas, dependientes del contexto y del juicio humano. Los resultados obtenidos en actividades claramente definidas no pueden extrapolarse automáticamente a todos los sectores de la economía.

A partir de las estimaciones existentes sobre exposición y mejoras de productividad, Acemoglu calcula que el aumento agregado de la productividad total de los factores podría ser significativo, pero moderado. Su estimación más amplia no supera el 0,66 % durante un período de diez años, mientras que una evaluación más prudente lo sitúa por debajo del 0,53 %. Estos cálculos no niegan que determinadas empresas, trabajadores o sectores puedan experimentar mejoras mayores, sino que cuestionan su traducción automática en un cambio macroeconómico generalizado.

El artículo introduce, además, una cuestión fundamental: las mejoras productivas no determinan por sí mismas cómo se distribuyen los beneficios. La IA puede elevar la productividad de trabajadores con menor cualificación en algunas tareas y, al mismo tiempo, aumentar la desigualdad si los beneficios son apropiados principalmente por las empresas o los propietarios del capital. Tampoco existe garantía de que una exposición relativamente amplia a la IA reduzca las diferencias salariales.

Acemoglu prevé que la IA puede ampliar la distancia entre las rentas del capital y las rentas del trabajo. Esta posibilidad refuerza la necesidad de diferenciar entre

9. ACEMOGLU, D., "The Simple Macroeconomics of AI", *Economic Policy*, vol. 40, núm. 121, 2025, pp. 13-58.

Si quieres adquirir esta obra haz click aquí

productividad agregada y prosperidad compartida. Una economía puede producir más valor sin que ese incremento se traduzca proporcionalmente en mejores salarios, mayor seguridad laboral o servicios públicos de mayor calidad.

La evidencia empírica disponible muestra, además, que los efectos laborales de la IA son heterogéneos. Albanesi[10] *et al.*, mediante el análisis de dieciséis países europeos entre 2011 y 2019, observan que la participación en el empleo aumentó, en promedio, en las ocupaciones más expuestas a la IA. Este incremento fue especialmente visible en ocupaciones con mayor presencia de trabajadores jóvenes y cualificados. Los autores encuentran también diferencias entre países vinculadas a la difusión tecnológica, la educación, la competencia y la protección del empleo.

Por el contrario, Bonfiglioli *et al.*[11] (2025), estudiando áreas laborales de Estados Unidos entre 2000 y 2020, identifican efectos negativos de la exposición a la IA sobre el empleo. Estos efectos recaen especialmente sobre trabajadores de baja cualificación y ocupaciones de producción, mientras que son más favorables para los trabajadores situados en la parte alta de la distribución salarial y para las profesiones STEM.

Los resultados no deben interpretarse necesariamente como incompatibles. Muestran que los efectos de la IA dependen del período analizado, la estructura productiva, las instituciones laborales, la cualificación y la forma en que la tecnología se adopta. La exposición tecnológica no produce consecuencias uniformes, sino que interactúa con las condiciones económicas y sociales de cada territorio.

Deranty y Corbin[12] subrayan igualmente que el impacto de la IA sobre el trabajo no puede reducirse a calcular el número de puestos susceptibles de automatización. Los trabajos no son conjuntos homogéneos de tareas separables. Incluyen conocimiento tácito, cooperación, relaciones humanas, juicio contextual y responsabilidad profesional.

La introducción de IA puede eliminar algunas tareas, transformar otras y crear nuevas actividades. También puede alterar la calidad del trabajo, la autonomía, las posibilidades de aprendizaje y la distribución del conocimiento dentro de la organización. Una evaluación completa debe estudiar no solo cuántos empleos desaparecen o se crean, sino qué tipo de trabajo se genera y quién controla el proceso productivo.

---

10. ALBANESI, S., DIAS DA SILVA, A., JIMENO, J. F., LAMO, A. y WABITSCH, A., "New Technologies and Jobs in Europe", *Economic Policy*, vol. 40, núm. 121, 2025, pp. 71-139.
11. BONFIGLIOLI, A., CRINÒ, R., GANCIA, G. y PAPADAKIS, I., "Artificial Intelligence and Jobs: Evidence from US Commuting Zones", *Economic Policy*, vol. 40, núm. 121, 2025, pp. 145-194.
12. DERANTY, J.-P. y CORBIN, T., "Artificial Intelligence and Work: A Critical Review of Recent Research from the Social Sciences", *AI & Society*, vol. 39, 2024, pp. 675-691.

Si quieres adquirir esta obra haz click aquí

La IA puede complementar el juicio profesional, facilitar el acceso a conocimientos especializados y liberar tiempo dedicado a actividades rutinarias. Sin embargo, también puede emplearse para estandarizar decisiones, intensificar la supervisión y reducir la capacidad de intervención de los trabajadores. La diferencia entre ambos resultados depende del diseño de la tecnología y de las decisiones organizativas que acompañan a su adopción.

4.4. DESIGUALDAD, DAÑOS SOCIOTÉCNICOS Y GRUPOS AFECTADOS

Los efectos sociales de la IA no se limitan al empleo y la renta. Los sistemas algorítmicos pueden afectar a la representación, la autonomía, la privacidad, el acceso a oportunidades y la participación social. Estos daños no siempre proceden de un error técnico claramente identificable. Pueden surgir de la interacción entre el modelo y una estructura social desigual.

Shelby *et al.*[13] proponen una taxonomía de daños sociotécnicos que permite ampliar el análisis más allá de la discriminación estadística. Los autores distinguen entre daños representacionales, de asignación, de calidad del servicio, interpersonales y sociales o sistémicos.

Los daños representacionales aparecen cuando un sistema reproduce estereotipos, invisibiliza a determinados grupos o los representa de forma degradante. Los daños de asignación se producen cuando la IA distribuye de manera desigual recursos, oportunidades o cargas. Los daños de calidad del servicio surgen cuando el sistema funciona peor para determinadas personas o colectivos.

Los daños interpersonales afectan a la autonomía, la privacidad, la identidad o las relaciones entre personas. Finalmente, los daños sociales o sistémicos comprenden efectos acumulativos sobre comunidades, instituciones y estructuras de poder. Esta última categoría resulta particularmente importante porque algunos perjuicios no son visibles al evaluar una única decisión, sino que aparecen mediante la repetición y generalización de prácticas algorítmicas.

La perspectiva sociotécnica muestra que un sistema puede alcanzar altos niveles de precisión y continuar produciendo consecuencias injustas. La precisión media puede ocultar diferencias entre grupos, mientras que una decisión técnicamente correcta puede reforzar una estructura social perjudicial.

13. SHELBY, R., RISMANI, S., HENNE, K., MOON, A., ROSTAMZADEH, N., NICHOLAS, P., YILLA-AKBARI, N. M., GALLEGOS, J., SMART, A., GARCIA, E. y VIRK, G., "Socio-technical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction", en *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, Association for Computing Machinery, 2023, pp. 723-741.

Si quieres adquirir esta obra haz click aquí

También deben considerarse los circuitos de retroalimentación. Las decisiones algorítmicas pueden modificar la realidad social que posteriormente se emplea para producir nuevos datos. Cuando una predicción condiciona la asignación de recursos, la vigilancia o las oportunidades, sus resultados pueden confirmar los supuestos iniciales del sistema y hacer que las desigualdades parezcan naturales o inevitables.

La evaluación debe incluir por ello los efectos colectivos, acumulativos y temporales, además de los resultados individuales. También requiere incorporar la experiencia de las personas y comunidades afectadas, puesto que algunos daños no pueden identificarse mediante métricas técnicas aisladas.

4.5. CONCENTRACIÓN ECONÓMICA Y CONTROL DE LAS INFRAESTRUCTURAS DE IA

La estructura económica del sector de la IA constituye otro componente central de su impacto social. El desarrollo de modelos avanzados requiere grandes cantidades de capacidad computacional, datos, talento especializado y financiación. Estas condiciones generan economías de escala que pueden favorecer la concentración del mercado.

Korinek y Vipra[14] analizan cómo la estructura de costes de los modelos fundacionales puede producir tendencias hacia la concentración, la integración vertical y el control de recursos esenciales por parte de un número limitado de empresas. La disponibilidad de computación, datos e infraestructura de distribución puede convertirse en una barrera de entrada para nuevos competidores.

Esta concentración no es únicamente un problema de precios o eficiencia económica. Las empresas que controlan los modelos y plataformas pueden determinar las condiciones de acceso, los tipos de aplicaciones desarrolladas y las prioridades de investigación. También pueden influir sobre los estándares técnicos, las prácticas laborales y la información disponible para ciudadanos e instituciones.

La concentración de inteligencia tecnológica puede traducirse, por tanto, en concentración de poder económico y social. Cuando unas pocas entidades controlan infraestructuras utilizadas por empresas, administraciones y ciudadanos, aumenta la dependencia respecto de sus decisiones y disminuye la capacidad de orientar la tecnología hacia objetivos colectivos.

Esta cuestión conecta directamente con *Power and Progress*. Acemoglu y Johnson, autores que ya se citaron anteriormente, sostienen que la concentración del poder

14. KORINEK, A. y VIPRA, J., "Concentrating Intelligence: Scaling and Market Structure in Artificial Intelligence", *Economic Policy*, vol. 40, núm. 121, 2025, pp. 225-256.

Si quieres adquirir esta obra haz click aquí

de decisión condiciona la dirección del cambio tecnológico. Cuando los trabajadores, las comunidades y las instituciones públicas quedan excluidos del proceso, la innovación tiende a reflejar prioritariamente los intereses de quienes controlan el capital y la infraestructura.

Una gobernanza sociotécnica de la IA debe incluir, por ello, políticas de competencia, interoperabilidad, acceso a infraestructuras, transparencia de las cadenas de valor y desarrollo de capacidades públicas. La regulación de cada aplicación individual resulta insuficiente cuando la estructura general del mercado limita las alternativas disponibles.

4.6. SOSTENIBILIDAD, COHESIÓN SOCIAL Y FINALIDAD DE LA INNOVACIÓN

La evaluación social de la IA debe considerar también su contribución a objetivos colectivos. Malmio[15] vincula la IA con la dimensión social del desarrollo sostenible y destaca aspectos relacionados con la seguridad económica, la desigualdad, la polarización, los valores y la resiliencia de las instituciones.

La sostenibilidad no se limita al consumo energético de los sistemas. Comprende igualmente la distribución de sus costes y beneficios, las oportunidades de acceso, la calidad del empleo, la protección de los grupos vulnerables y la capacidad de las sociedades para controlar las tecnologías de las que dependen.

Una innovación puede ser rentable para una empresa y, al mismo tiempo, generar costes significativos para trabajadores, comunidades o instituciones públicas. La evaluación de la IA debe incorporar estas externalidades y evitar que sus beneficios privados se analicen de manera separada de sus consecuencias sociales.

Watson, Mökander y Floridi[16] proponen un pragmatismo sociotécnico que evita tanto el solucionismo tecnológico como el rechazo general de la IA. El primero supone que los problemas sociales pueden resolverse mediante sistemas técnicamente más avanzados. El segundo considera que la tecnología reproduce inevitablemente las estructuras de poder existentes y que no existe margen para orientar su desarrollo.

El pragmatismo sociotécnico reconoce que los efectos de la IA no están predeterminados. Dependen de las instituciones, las normas y la capacidad de los actores sociales para influir en su diseño y aplicación. Esta posición permite abordar

15. MALMIO, I., "Artificial Intelligence and the Social Dimension of Sustainable Development: Through a Security Perspective", *Discover Sustainability*, vol. 5, 2024, art. 466.
16. WATSON, D. S., MÖKANDER, J. y FLORIDI, L., "Competing Narratives in AI Ethics: A Defense of Sociotechnical Pragmatism", *AI & Society*, vol. 40, 2025, pp. 3163-3185.

Si quieres adquirir esta obra haz click aquí

los riesgos de forma crítica sin renunciar a emplear la tecnología para finalidades socialmente valiosas.

4.7. IMPLICACIONES PARA UNA INTELIGENCIA ARTIFICIAL CONFIABLE

La perspectiva sociotécnica y socioeconómica modifica el significado de la confiabilidad. Una IA confiable no es simplemente aquella que funciona correctamente desde un punto de vista técnico. Debe insertarse en un sistema institucional y económico capaz de orientar sus usos, distribuir de manera justa sus beneficios y prevenir sus efectos perjudiciales.

Un modelo robusto puede formar parte de un sistema poco confiable cuando se utiliza para una finalidad inadecuada, cuando reduce injustificadamente la autonomía profesional o cuando sus beneficios se concentran mientras sus costes recaen sobre grupos vulnerables. De forma similar, una aplicación puede cumplir requisitos técnicos específicos y contribuir al mismo tiempo a la concentración del mercado o al deterioro de las condiciones laborales.

La confiabilidad debe comprender, por tanto, al menos cinco dimensiones sociotécnicas:

1. Una finalidad socialmente legítima y claramente definida.
2. Una distribución justa de los beneficios, riesgos y oportunidades.
3. Una integración organizativa que preserve la capacidad y autonomía humanas.
4. Una participación efectiva de los trabajadores y grupos afectados.
5. Una supervisión institucional capaz de evaluar efectos individuales, colectivos y económicos.

La principal conclusión de esta sección es que el progreso tecnológico y el progreso social no son equivalentes. La IA puede aumentar determinadas capacidades productivas sin generar automáticamente una prosperidad compartida. Sus efectos dependen de quién orienta la innovación, qué problemas se decide resolver y qué instituciones participan en la distribución de sus beneficios.

En consecuencia, la pregunta no debe limitarse a cuánto crecimiento o productividad puede producir la IA. Debe incluir quién se beneficia, quién soporta los costes, qué tareas se automatizan, qué capacidades humanas se fortalecen y qué estructuras de poder se consolidan.

Si quieres adquirir esta obra haz click aquí

La construcción de una IA confiable exige, en definitiva, algo más que algoritmos precisos y seguros. Requiere orientar el cambio tecnológico hacia la ampliación de las capacidades humanas, la calidad del trabajo, la igualdad, la competencia económica y el bienestar colectivo. La IA no determina de manera inevitable el futuro de la sociedad: las decisiones sociales e institucionales determinan qué tipo de IA se desarrolla y a qué forma de progreso contribuye.

V. CONCLUSIONES

El desarrollo de una IA confiable exige integrar la reflexión ética y las garantías jurídicas. La ética permite examinar las finalidades, los valores y las responsabilidades que deben orientar la IA, mientras que el Derecho transforma parte de esas exigencias en obligaciones vinculantes. El AI Act y su evolución normativa en curso mediante el Digital Omnibus proporcionan un marco esencial para gestionar los riesgos, proteger los derechos fundamentales y distribuir responsabilidades, pero el cumplimiento normativo no puede sustituir la deliberación sobre la legitimidad y conveniencia de cada aplicación.

La IA debe analizarse, además, como un sistema sociotécnico inserto en organizaciones, mercados e instituciones. Sus efectos no dependen únicamente del modelo o de los datos, sino también de las decisiones humanas, los incentivos económicos y las relaciones de poder que condicionan su utilización. La discriminación, la transformación del empleo, la pérdida de autonomía profesional, la desigualdad, los daños colectivos y la concentración de capacidades tecnológicas muestran que la confiabilidad no es una propiedad exclusiva del algoritmo, sino del sistema social y organizativo en el que este se integra.

En consecuencia, avanzar hacia la confianza en la IA requiere una gobernanza capaz de articular ética, Derecho, garantías técnicas y evaluación sociotécnica durante todo el ciclo de vida. El objetivo no debe ser conseguir que las personas confíen indiscriminadamente en la tecnología, sino desarrollar sistemas, organizaciones e instituciones que sean merecedores de esa confianza. Solo una IA orientada hacia finalidades legítimas, sometida a responsabilidades efectivas y evaluada atendiendo a sus consecuencias humanas, sociales y económicas puede fundamentar una confianza crítica y justificada.

VI. BIBLIOGRAFÍA

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy, 40*(121), 13-58. doi:10.1093/epolic/eiae042

Si quieres adquirir esta obra haz click aquí

Si quieres adquirir esta obra haz click aquí



ESTUDIOS

¿Está el Derecho preparado para afrontar los desafíos de la inteligencia artificial? ¿Puede la innovación tecnológica avanzar sin comprometer los derechos fundamentales ni alterar categorías esenciales del Derecho? ¿Qué papel debe desempeñar la ética en la construcción de un marco regulatorio capaz de responder a una realidad en constante transformación?

Esta obra colectiva ofrece una respuesta rigurosa e interdisciplinar a estas cuestiones mediante un análisis integral de la relación entre la inteligencia artificial, el Derecho y la ética. Constituye una aportación singular en la doctrina española reciente, al abordar de forma conjunta la regulación jurídica, los fundamentos técnicos de la IA y sus implicaciones éticas, sociales y democráticas.

Desde una perspectiva jurídica, la obra examina cuestiones de máxima actualidad, como el Reglamento Europeo de Inteligencia Artificial, su aplicación institucional y extraterritorial, así como su incidencia en diversas ramas del ordenamiento, entre ellas el Derecho constitucional, penal, laboral, civil, registral y tributario. Asimismo, analiza los mecanismos de protección y reclamación de las personas afectadas frente a los riesgos derivados del uso de sistemas de IA.

El estudio se inicia con los fundamentos de la IA responsable, confiable y explicable, y se completa con capítulos dedicados a aspectos esenciales de la tecnología, como los modelos de IA de propósito general, el tratamiento del lenguaje, la IA generativa y agéntica, la seguridad y la sostenibilidad. Junto a ello, la dimensión ética ocupa un lugar central, abordándose cuestiones como los sesgos humanos y algorítmicos, los riesgos para la democracia y la autonomía personal, la gobernanza de la IA y su impacto social y económico.

Dirigida a investigadores, profesionales del Derecho y de la tecnología, responsables públicos y estudiantes, esta obra constituye una referencia indispensable para comprender uno de los mayores retos jurídicos, tecnológicos y éticos de nuestro tiempo.

ISBN: 978-84-1078-367-6



AENOR GESTIÓN DE LA CALIDAD ISO 9001 ER-0280/2005
IQNET RECOGNIZED CERTIFICATION
AENOR GESTIÓN AMBIENTAL ISO 14001 GA-2005/0100

UNIVERSIDAD DE GRANADA

DASCI
Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional

ARANZADI