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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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