DERECHO DIGITAL Y PROPIEDAD INTELECTUAL

7

INTELIGENCIA ARTIFICIAL Y DERECHO DE DAÑOS: CUESTIONES ACTUALES

Acorde al Reglamento (UE) 2024/1689

Itziar Alkorta Idiakez
Cristina Argelich Comelles
Maria Cristina Berenguer Albaladejo
Yolanda Bustos Moreno
Maria Raquel Evangelio Llorca
Beatriz Extremera Fernández
Pedro José Femenía López
María Remedios Guilabert Vidal
María Jorqui Azofra
Raúl Lafuente Sánchez
Pedro José López Mas
Raquel Luquin Bergareche
Andrés Marín Salmerón
Luz Martínez Velencoso
Lucía Molina Martínez
Óscar Monje Balmaseda
Esther Monterroso Casado
Juan Antonio Moreno Martínez
Carmen Muñoz García
Alberto Muñoz Villarreal
Íñigo Navarro Mendizábal
Manuel Ortiz Fernández
Miquel Peguera Poch
Antonio Rubí Puig
Alberto Tapia Hermida

Dykinson, S.L.

MORENO MARTÍNEZ, J.A.
FEMENÍA LÓPEZ, P.J.
(Coordinadores)

INTELIGENCIA ARTIFICIAL
Y DERECHO DE DAÑOS:
CUESTIONES ACTUALES

Acorde al Reglamento (UE) 2024/1689

COLECCIÓN
DERECHO DIGITAL Y PROPIEDAD INTELECTUAL

INTELIGENCIA ARTIFICIAL Y DERECHO DE DAÑOS: CUESTIONES ACTUALES

Acorde al Reglamento (UE) 2024/1689

Moreno Martínez, J.a.
Femenía López, P.j.
(Coordinadores)

Itziar Alkorta Idiakez
Cristina Argelich Comelles
Maria Cristina Berenguer Albaladejo
Yolanda Bustos Moreno
Maria Raquel Evangelio Llorca
Beatriz Extremera Fernández
Pedro José Femenía López
María Remedios Guilabert Vidal
María Jorqui Azofra
Raúl Lafuente Sánchez
Pedro José López Mas
Raquel Luquin Bergareche
Andrés Marín Salmerón
Luz Martínez Velencoso
Lucía Molina Martínez
Óscar Monje Balmaseda
Esther Monterroso Casado
Juan Antonio Moreno Martínez
Carmen Muñoz García
Alberto Muñoz Villarreal
Íñigo Navarro Mendizábal
Manuel Ortiz Fernández
Miquel Peguera Poch
Antonio Rubí Puig
Alberto Tapia Hermida

Dykinson, S.L.

Este libro ha sido sometido a evaluación por parte de nuestro Consejo Editorial.
Para mayor información, véase *www.dykinson.com/quienes_somos*

Este trabajo se enmarca en el Proyecto I+D+i (Referencia: PID2020-116185GB-100) del Ministerio de Ciencia e Innovación: "La irrupción de la inteligencia artificial en el Derecho de Daños y su adaptación a las nuevas tecnologías", siendo investigadores principales los profesores Juan Antonio Moreno Martínez y Pedro José Femenía López.

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

ISBN: 978-84-1070-708-5
Depósito Legal: M-25437-2024
DOI: https://doi.org/10.14679/3532

ISBN electrónico: 978-84-1122-801-5

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

Índice

La discriminación algorítmica en el sector sanitario 1

Itziar Alkorta Idiakez

1. INTRODUCCIÓN 1
2. CASOS DE DISCRIMINACIÓN ALGORÍTMICA EN EL SECTOR SANITARIO 3
3. APLICABILIDAD LA NORMATIVA ANTIDISCRIMINATORIA EN MATERIA DE DISCRIMINACIÓN ALGORÍTMICA 6
   - **3.1. Normativa antidiscriminatoria** 7
   - **3.2. Limitaciones de la eficacia horizontal** 9
   - **3.3. La prueba del daño moral** 10
   - **3.4. Litigación colectiva** 13
4. APLICABILIDAD DE LA NORMATIVA SECTORIAL DE LA IA 15
   - **4.1. Principios y requisitos aplicables a la seguridad de los productos sanitarios con IA** 15
   - **4.2. La falta de transparencia en las decisiones automatizadas** 17
   - **4.3. El problema de la calidad de los conjuntos de datos** 20
   - **4.4. La responsabilidad por daños morales causados por la IA** 24
5. CONCLUSIONES 26

La armonización del tratamiento legal de la responsabilidad civil contractual y extracontractual del metaverso con la regulación europea sobre plataformas en línea 31

Cristina Argelich Comelles

1. CONSIDERACIONES INICIALES ACERCA DEL METAVERSO Y LA RESPONSABILIDAD CIVIL 31
2. IDENTIDAD DIGITAL DEL RESPONSABLE CIVIL Y PROPIEDAD DE LOS ACTIVOS DIGITALES PATRIMONIALES 33

3. EL RÉGIMEN DE RESPONSABILIDAD DEL PROVEEDOR DE SERVICIOS DE LA PLATAFORMA Y DEL USUARIO PROFESIONAL EN EL ORDENAMIENTO JURÍDICO EUROPEO 35

3.1. La incardinación del régimen jurídico de las plataformas en línea en la responsabilidad civil contractual: hacia un sistema de responsabilidad civil objetiva por pérdida o desprogramación de un activo digital y por discriminación algorítmica 39

3.2. La incardinación del régimen jurídico de las plataformas en línea en la responsabilidad extracontractual por los daños causados en las plataformas del Metaverso 43

4. REFLEXIONES PROSPECTIVAS SOBRE LA RESPONSABILIDAD CIVIL CONTRACTUAL Y EXTRACONTRACTUAL: EL INFORME ESPAÑOL PARA LA COMISIÓN EUROPEA EN MATERIA DE CONTRATACIÓN CON INTELIGENCIA ARTIFICIAL 44

BIBLIOGRAFÍA 46

Transparencia y explicabilidad para prevenir la discriminación de los sistemas de inteligencia artificial: la interacción entre el RGPD y el RIA 49

Mª Cristina Berenguer Albaladejo

1. LA DISCRIMINACIÓN ALGORÍTMICA COMO UNO DE LOS PRINCIPALES RIESGOS DERIVADOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL PARA LA TOMA DE DECISIONES 50

2. LA OPACIDAD COMO PRINCIPAL ESCOLLO PARA DETECTAR Y DEMOSTRAR LA DISCRIMINACIÓN ALGORÍTMICA 55

2.1. Consideraciones previas 55

2.2. Opacidad en el uso y sobre el contenido de los algoritmos 57

2.3. Opacidad jurídica y técnica del algoritmo 59

3. TRANSPARENCIA ALGORÍTMICA Y EXPLICABILIDAD: ¿QUÉ IMPLICAN ESTAS EXIGENCIAS? 68

4. MEDIDAS PARA GARANTIZAR LA TRANSPARENCIA Y LA EXPLICABILIDAD EN LA TOMA DE DECISIONES ALGORÍTMICAS 75

4.1 Estado de la cuestión 75

4.2 La transparencia y la explicabilidad en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, de protección de datos (RGPD): especial referencia a las decisiones automatizadas del art. 22 78

4.3. La transparencia y la explicabilidad en el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial 101

5. CONSIDERACIONES FINALES SOBRE LA NECESIDAD DE TRANSPARENCIA Y EXPLICABILIDAD PARA DETECTAR Y DEMOSTRAR LA DISCRIMINACIÓN ALGORÍTMICA 112
BIBLIOGRAFÍA 113

Aplicaciones de la inteligencia artificial conforme a la Ley de Movilidad Sostenible. Consideraciones en torno al régimen de responsabilidad civil acorde con la innovación 119

Yolanda Bustos Moreno

1. EL REGLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO Y DEL CONSEJO DE 13 DE JUNIO DE 2024 POR EL QUE SE ESTABLECEN NORMAS ARMONIZADAS EN MATERIA DE INTELIGENCIA ARTIFICIAL Y EL PROYECTO DE LEY DE MOVILIDAD SOSTENIBLE DE 23 DE FEBRERO DE 2024 120
   - **1.1. Consideraciones generales de la AIA** 120
   - **1.2. La regulación y su papel de apoyo a la innovación en el desarrollo de sistemas de IA** 122
   - **1.3. El Proyecto de Ley de Movilidad Sostenible de 23 de febrero de 2024 con relación a la aplicación de la IA en vehículos automatizados** 124
   - **1.4. El concepto de "sistema de inteligencia artificial" en la AIA y PLMS** 126
2. DILEMAS EN TORNO A LA REGULACIÓN DE LA RESPONSABILIDAD CIVIL EN LAS ACTIVIDADES QUE EMPLEAN SISTEMAS DE IA . 129
   - **2.1. Características especiales de los sistemas de IA con relación al riesgo** 130
   - **2.2. El debate sobre el régimen de responsabilidad civil más favorable a la innovación en sistemas de IA** 137
   - **2.3. El replanteamiento de la responsabilidad objetiva en el *Complementary Impact Assessment. Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence*** 139
3. EL APOYO A LOS SISTEMAS DE IA INNOVADORES ANTES DE LA INTRODUCCIÓN EN EL MERCADO O PUESTA EN SERVICIO DESDE EL PERFIL DE LA RESPONSABILIDAD CIVIL. 141

BIBLIOGRAFÍA 145

Responsabilidad civil e inteligencia artificial en el ámbito sanitario: posibles vías de reclamación 149

Raquel Evangelio Llorca

1 APLICACIONES DE LA INTELIGENCIA ARTIFICIAL EN EL SECTOR SANITARIO 150
2. RESPONSABILIDAD CIVIL POR DAÑOS CAUSADOS POR EL USO DE SISTEMAS DE INTELIGENCIA DE ARTIFICIAL EN EL ÁMBITO DE LA SANIDAD: CUESTIONES GENERALES 155
3. DAÑOS CAUSADOS POR LA INTELIGENCIA ARTIFICIAL COMO PRODUCTO DEFECTUOSO 166
3.1. Ámbito de aplicación del régimen de responsabilidad civil por daños causados por productos defectuosos. Los sistemas inteligentes como productos defectuosos 166
3.2. Sujetos responsables 178
3.3. Sujetos legitimados para ejercitar acciones por daños causados por productos defectuosos 186
3.4. Fundamento de la responsabilidad y causas de exoneración 187
4. RÉGIMEN DE RESPONSABILIDAD CIVIL POR DAÑOS CAUSADOS POR SERVICIOS SANITARIOS DEL ART. 148 TRLGDCU 190
4.1. Ámbito de aplicación y fundamento de la responsabilidad 190
4.2. Sujeto responsable 195
4.3. Sujeto protegido 197
5. RESPONSABILIDAD PATRIMONIAL DE LA ADMINISTRACIÓN SANITARIA 199
6. RÉGIMEN DE RESPONSABILIDAD CIVIL EXTRACONTRACTUAL DEL CÓDIGO CIVIL 204
7. CONSIDERACIONES FINALES SOBRE LA CONCURRENCIA DE REGÍMENES APLICABLES 210
8. BIBLIOGRAFÍA 214

Los *deepfakes* y la intromisión en los derechos de la personalidad (imagen, voz, honor y protección de datos) y sus mecanismos de reparación 223

Beatriz Extremera Fernández

1. INTRODUCCIÓN 223
2. PRECISIONES CONCEPTUALES: QUÉ ES EL *DEEPFAKE* Y SU CLASIFICACIÓN DEL RIESGO 225
3. PROBLEMÁTICA JURÍDICA DEL *DEEPFAKE* 230

3.1. Los derechos al honor, a la propia imagen y a la voz en la LO 1/1982 230
3.2. La imagen y voz como datos de carácter personal en el uso del *deepfake* 243
4. EL PAPEL DE LA ADVERTENCIA EN EL USO DEL *DEEPFAKE* 246
5. MECANISMOS DE PROTECCIÓN 248
5.1. Tutela de los derechos de la personalidad protegidos en la LO 1/1982 249
5.2. Tutela de los datos de carácter personal 250
5.3. La responsabilidad de los prestadores de servicios de la sociedad digital 253
6. CONCLUSIONES 255
7. BIBLIOGRAFÍA. 257

Responsabilidad civil derivada de la adquisición y utilización de *werables* y servicios digitales en materia de salud 261

Pedro J. Femenía López.

1. PLANTEAMIENTO: DE LA *E-HEALTH* A LA AUTONOMÍA INDIVIDUAL EN LA GESTIÓN DE LA SALUD 261
2. RESPONSABILIDAD DERIVADA DE LA COMPRA DEL BIEN O DE LA CONTRATACIÓN DEL CONTENIDO O SERVICIO 269
2.1. Ámbito de aplicación 269
2.2. Sujeto responsable 274
2.3. Criterios de imputación 275
3. LA RESPONSABILIDAD CIVIL DERIVADA DEL USO DE *WEREABLES* Y SERVICIOS DIGITALES EN MATERIA DE SALUD 281
3.1. Ámbito de aplicación 283
3.2. Sujetos responsables 293
3.3. Criterios de imputación 300
BIBLIOGRAFÍA 315

Interfaces cerebro-computador: protección de los neurodatos a través de los neuroderechos y de la responsabilidad civil del art. 82 del RGPD 319

María Remedios Guilabert Vidal

1. INTRODUCCIÓN 319
1.1. El estado actual de la Neurotecnología: avances y desafíos 319

1.2. Las interfaces cerebro-computador 325
2. LA PROTECCIÓN DISPENSADA POR LOS NEURODERECHOS 329
2.1. Los neuroderechos como nuevos derechos fundamentales: concepto y clases 329
2.2. *Soft law* público y avances legislativos 331
3. PROTECCIÓN DISPENSADA A LOS NEURODATOS POR EL REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO 336
3.1. Concepto y naturaleza jurídica del neurodato 336
3.2. Responsabilidad por daños causados por infracción del derecho a la protección de datos en el ámbito de las BCI 338
BIBLIOGRAFÍA 349

Encaje del sistema de Inteligencia Artificial utilizado con determinados fines médicos en algunas de las cuestiones suscitadas al amparo del régimen de responsabilidad por productos defectuosos 353

María Jorqui Azofra

1. INTRODUCCIÓN 353
2. EL SISTEMA DE IA COMO PRODUCTO 356
3. EL SISTEMA DE IA COMO PRODUCTO SANITARIO 360
4. ¿QUÉ DETERMINA EL CARÁCTER DEFECTUOSO DEL SISTEMA DE IA? 365
5. SISTEMA DE EXHIBICIÓN DE PRUEBAS Y CARGA DE LA PRUEBA 380
6. CAUSAS DE EXONERACIÓN: ESPECIAL CONSIDERACIÓN A LOS RIESGOS DEL DESARROLLO 385
7. CONCLUSIONES 390
BIBLIOGRAFÍA 393
NORMATIVA Y OTROS DOCUMENTOS 396
JURISPRUDENCIA 396

IA y vehículos autónomos: cuestiones concernientes a la responsabilidad no contractual en la vertiente del derecho internacional privado 399

Raúl Lafuente Sánchez

1. INTRODUCCIÓN 400
2. VEHÍCULOS AUTÓNOMOS Y RESPONSABILIDAD CIVIL EXTRACONTRACTUAL 403

2.1 Incidencia del Reglamento de Inteligencia Artificial 403
2.2 Propuesta de revisión de la Directiva 85/374 sobre productos defectuosos 407
3. SOLUCIÓN DE CONTROVERSIAS Y APLICACIÓN DE LAS NORMAS DE DERECHO INTERNACIONAL PRIVADO 415
3.1 Competencia judicial internacional 415
3.2 Ley aplicable 423
4. REFLEXIONES FINALES: IDONEIDAD DE LOS INSTRUMENTOS DE DIPR ACTUALMENTE EN VIGOR PARA REGULAR LAS RECLAMACIONES DERIVADAS DE LA CONDUCCIÓN AUTOMATIZADA 444
4.1 Para determinar la jurisdicción de los tribunales de la UE 444
4.2 En materia de ley aplicable 445
BILIOGRAFÍA 446

Vehículos autónomos y responsabilidad civil. La vacilante ruta marcada por el legislador europeo 451

Pedro José López Mas

1. CONSIDERACIONES PRELIMINARES SOBRE LA CONDUCCIÓN AUTOMATIZADA 452
1.1. Conceptualización y situación actual 452
1.2. Retos jurídicos que presenta este «novedoso» fenómeno 456
2. RÉGIMEN JURÍDICO DE LA RESPONSABILIDAD CIVIL DERIVADA DEL USO DE VEHÍCULOS A MOTOR, Y BREVES NOTAS SOBRE SU ASEGURAMIENTO 459
2.1. Planteamiento de la cuestión 459
2.2. El concepto de «vehículo a motor» 463
2.3. El concepto de «hecho de la circulación» 467
2.4. El concepto de «conductor» 469
3. LA INCIDENCIA EN LA CONDUCCIÓN AUTOMATIZADA DE LA NUEVA PROPUESTA DE DIRECTIVA SOBRE RESPONSABILIDAD CIVIL EN MATERIA DE INTELIGENCIA ARTIFICIAL, Y SUS EVIDENTES DISFUNCIONALIDADES 470
3.1. Ámbito de aplicación y caracteres 473
3.2. Deber de exhibición de pruebas y presunción *iuris tantum* en caso de incumplimiento 475
3.3. Presunción *iuris tantum* de la relación de causalidad en caso de culpa 476
4. BIBLIOGRAFÍA 479

Inteligencia artificial en la prestación de servicios de salud: funcionalidades, riesgos y responsabilidad civil 481

Raquel Luquin Bergareche

1. INTRODUCCION. ROBOTS Y APLICACIONES DE INTELIGENCIA ARTIFICIAL COMO INSTRUMENTOS AUXILIARES EN LA PRESTACION DE SERVICIOS MEDICOS 482
2. LA PREVENCION DE LOS RIESGOS DE LA INTELIGENCIA ARTIFICIAL EN SALUD A LA LUZ DEL REGLAMENTO (UE) 2024/1689 DE 13 DE JUNIO DE 2024, POR EL QUE SE ESTABLECEN NORMAS ARMONIZADAS EN MATERIA DE IA (RIA) 491
   - **2.1. Primer marco regulatorio europeo de la IA** 491
   - **2.2. Riesgos y salud: la ambigüa definición de los sistemas IA de alto riesgo** 493
   - **2.3. Obligaciones de proveedores y responsables del despliegue: información y supervisión** 500
   - **2.4. Aplicaciones de IA en salud para uso particular o doméstico** 506
   - **2.5. El RIA como sistema normativo de prevención del riesgo: remisión a otros marcos regulatorios en el ámbito de los daños causados por sistemas de IA en salud** 509
   - **2.6. Formación y capacitación en IA del profesional de la salud** 512
3. DAÑOS CAUSADOS EN INTERVENCIONES MEDICAS CON AUXILIO DE IA: REDEFINICION DE LA "LEX ARTIS" Y FUNDAMENTOS DE LA RESPONSABILIDAD 513
   - **3.1. Cuando el médico se prevale de un sistema de IA y su actuación causa daños: presupuestos de la obligación de responder** 513
   - **3.2. Caracteres de los sistemas de IA en salud: en particular, la influencia del grado de autonomía del robot o sistema auxiliar de IA en la responsabilidad por daños** 518
   - **3.3. Relación de causalidad. La causalidad física y su prueba** 521
   - **3.4. La causalidad jurídica: el juicio de imputación** 523
   - **3.5. Agentes implicados en la prestación de servicios médicos con auxilio de IA** 524
   - **3.6. Causas de exclusión o exoneración** 529
4. ALGUNAS REFLEXIONES SOBRE EL RÉGIMEN (NO ARMONIZADO Y "DE MÍNIMOS") DE LA PROPUESTA DE DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO RELATIVA A LA ADAPTACIÓN DE LAS NORMAS DE RESPONSABILIDAD CIVIL EXTRACONTRACTUAL A LA IA (PDRCIA) 531
5. REFERENCIAS BIBLIOGRAFICAS 533

La doctrina *crashworthiness*: origen, desarrollo y posible aplicación a los vehículos automatizados 539

Andrés Marín Salmerón

1. LA DOCTRINA *CRASHWORTHINESS* O *SECOND COLLISION* 540
   - **1.1. Breve referencia a su concepto y objetivo del trabajo** 540
   - **1.2. Principios y orígenes de la doctrina *crashworthiness*** 544
   - **1.3. Aplicación de la doctrina *Crashworthiness*. Relación de la primera colisión con la *second collision:* intervención de tercero y culpa del perjudicado** 555
2. SU CONEXIÓN CON EL CRITERIO DE RIESGO UTILIDAD Y EL DISEÑO ALTERNATIVO RAZONABLE: DE NUEVO CON LA RESPONSABILIDAD SUBJETIVA 567
3. LA DOCTRINA *CRASHWORTHINESS* EN LA JURISPRUDENCIA ESPAÑOLA 569
4. LA APLICACIÓN DE LA DOCTRINA EN ESPAÑA: SU COMPATIBILIDAD CON EL REAL DECRETO LEGISLATIVO 8/2004, DE 29 DE OCTUBRE, POR EL QUE SE APRUEBA EL TEXTO REFUNDIDO DE LA LEY SOBRE RESPONSABILIDAD CIVIL Y SEGURO EN LA CIRCULACIÓN DE VEHÍCULOS A MOTOR 573
5. LA APLICACIÓN DE LA DOCTRINA *CRASHWORTHINESS* CON LA NUEVA NORMATIVA DE RESPONSABILIDAD POR DAÑOS POR PRODUCTOS DEFECTUOSOS 577
6. BIBLIOGRAFÍA 579

El uso de algoritmos en detrimento de los principios jurídicos y económicos de la Unión Europea 583

Luz M. Martínez Velencoso

1. INTRODUCCIÓN 583
2. TRANSPARENCIA ALGORÍTMICA 585
   - **2.1. Derecho de la competencia** 585
   - **2.2. Transparencia en la publicidad algorítmica** 593
3. DERECHO DE CONSUMO E INTELIGENCIA ARTIFICIAL 596
   - **3.1. Microtargeting** 596
   - **3.2. Contratos algorítmicos** 599
4. BIBLIOGRAFÍA 600

Uso de inteligencia artificial, *Big Data* y otras tecnologías disruptivas en las plataformas digitales de alojamiento turístico: desafíos actuales en materia de privacidad, transparencia algorítmica y responsabilidad civil 603

Lucía Molina Martínez

1. *BIG DATA*, INTELIGENCIA ARTIFICIAL, IoT Y TECNOLOGÍA *BLOCKCHAIN* EN LAS PLATAFORMAS DIGITALES DE ALOJAMIENTO TURÍSTICO 604
   - **1.1. La transformación digital del sector turístico: el papel de las plataformas digitales de alojamiento turístico** 604
   - **1.2. La aplicación de tecnologías innovadoras disruptivas por las plataformas de alojamiento turístico: desde el algoritmo hasta la tecnología *blockchain*** 607
2. IMPACTO DE LAS TECNOLOGÍAS DISRUPTIVAS EN LA PRIVACIDAD Y PROTECCIÓN DE DATOS DE LAS PLATAFORMAS DE ALOJAMIENTO TURÍSTICO 613
   - **2.1. Empleo de tecnologías disruptivas en la recopilación y tratamiento masivo de datos personales: aparición de nuevas categorías de datos y riesgos para la privacidad de los usuarios** 613
   - **2.2. La elaboración de perfiles y la adopción de decisiones automatizadas a través de sistemas avanzados de IA** 620
3. TRANSPARENCIA ALGORÍTMICA Y RESPONSABILIDAD CIVIL EN EL MARCO DE LA INTERMEDIACIÓN DE LAS PLATAFORMAS DE ALOJAMIENTO TURÍSTICO 628
   - **3.1. Desafíos que plantea la toma de decisiones algorítmicas y la regulación europea en materia de IA para combatirlos** 628
   - **3.2. Exigencias de transparencia para los sistemas algorítmicos de recomendación, clasificación, selección de contenidos y publicidad en línea de los prestadores de servicios de alojamiento de datos** 632
   - **3.3. Tratamiento legal de la responsabilidad de las plataformas por la moderación automatizada de contenidos y el incumplimiento de las obligaciones de transparencia algorítmica: régimen transitorio a la espera de una regulación específica acerca de la discriminación algorítmica** 640

BIBLIOGRAFÍA 645

Implicaciones jurídicas del uso de los robots y la inteligencia artificial en el ámbito sanitario. ¿Hacia una nueva medicina? 651

Óscar Monje Balmaseda

1. LA PROTECCIÓN DE LA SALUD Y LA EVOLUCIÓN TECNOLÓGICA: ESPECIAL REFERENCIA A LA ROBÓTICA Y LA INTELIGENCIA ARTIFICIAL 651
   - **1.1. Consideraciones previas: la robótica y la inteligencia artificial en el ámbito sanitario** 651
   - **1.2. La utilización de la inteligencia artificial en el ámbito de la salud: sus limitaciones y los desafíos éticos y jurídicos que presenta**. 654
2. PLANTEAMIENTO LEGISLATIVO EN MATERIA DE INTELIGENCIA ARTIFICIAL Y RESPONSABILIDAD CIVIL EN LA UNIÓN EUROPEA 660
   - **2.1. La responsabilidad civil en el ámbito sanitario. Responsabilidad objetiva y gestión de riesgos** 660
   - **2.2. El posicionamiento inicial de la Unión Europea en materia de responsabilidad civil de los robots y los sistemas de inteligencia artificial** 664
   - **2.3. Las propuestas de regulación de la UE: La Directiva sobre responsabilidad por daños causados por productos defectuosos y la Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial** 672

BIBLIOGRAFÍA UTILIZADA 679

La responsabilidad civil derivada de los accidentes de circulación ocasionados con vehículos autónomos 681

Esther Monterroso Casado

1. INTRODUCCIÓN 682
2. EVOLUCIÓN Y REGULACIÓN DE LA RESPONSABILIDAD CIVIL EXTRACONTRACTUAL POR DAÑOS EN LA CIRCULACIÓN DE VEHÍCULOS A MOTOR 683
   - **2.1. Evolución legal de la responsabilidad derivada de los accidentes de circulación** 683
   - **2.2. Regulación actual y perspectivas de futuro de la responsabilidad derivada de los accidentes de circulación** 687
3. VEHÍCULOS AUTÓNOMOS Y CONDUCCIÓN AUTOMATIZADA 692
   - **3.1. El vehículo autónomo** 692
   - **3.2. Los niveles de autonomía** 694
   - **3.3. Autonomía real en la oferta de conducción automatizada** 696

4. REGULACIÓN DE LA CONDUCCIÓN AUTOMATIZADA 698
4.1. Marco jurídico europeo de vehículos automatizados y totalmente automatizados 698
4.2. Marco jurídico nacional de conducción automatizada 703
5. REGULACIÓN DE LOS SISTEMAS DE ALTO RIESGO EN LA INTELIGENCIA ARTIFICIAL 712
5.1. Reglamento europeo por el que se establecen normas armonizadas en materia de inteligencia artificial 712
5.2. Directiva sobre responsabilidad por los daños causados por productos defectuosos 717
5.3. Propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial 720
6. HACIA UN NUEVO CRITERIO DE RESARCIMIENTO DE DAÑOS DERIVADO DE LA AUSENCIA DEL CONDUCTOR DEL VEHÍCULO ... 726
6.1. Responsabilidad del fabricante del vehículo 729
6.2. Responsabilidad del operador o del propietario del vehículo 732
6.3. Resarcimiento del daño por la aseguradora del vehículo, tomando como referencia la LRCSCVM 734
6.4. Resarcimiento del daño por la aseguradora del vehículo, sin imputación de la responsabilidad 737
7. CONCLUSIONES 739
8. BIBLIOGRAFÍA 743

Impresión 3D en el ámbito médico: problemática de la responsabilidad civil -y patrimonial- y sus incidencias digitales y de inteligencia artificial por las reformas de la Unión Europea 749

Juan Antonio Moreno Martínez

1. LA FABRICACIÓN ADITIVA O IMPRESIÓN EN 3D: LAS INICIATIVAS DE LA UNIÓN EUROPEA 750
2. LA BIOIMPRESIÓN 3D COMO ESPECÍFICA IMPRESIÓN EN LA MEDICINA. LA RESPONSABILIDAD CIVIL -Y PATRIMONIAL-: RÉGIMEN LEGAL APLICABLE 755
2.1. Consideraciones generales 755
2.2. Incidencia de la consideración de la bioimpresión como producto sanitario: Evaluación de la conformidad. La responsabilidad patrimonial de la Agencia Española del medicamento y productos sanitarios (AEMPS) y su delimitación con respecto a los casos de responsabilidad patrimonial de la Administración sanitaria 760

2.3. Responsabilidad civil en la bioimpresión 767
BIBLIOGRAFÍA 782

Taxonomía de los modelos de IA de uso general. Probabilidad de generar riesgos de alto impacto y la necesidad de identificarlos 787

Carmen Muñoz García

1. JUSTIFICACIÓN DEL ESTUDIO 787
   1.1. La IA Generativa como modelo de IA de uso general. El caso 787
   1.2. ¿Por qué regularlo? 790
   1.3. La incidencia en los derechos de la persona 793
2. TAXONOMÍA DE LOS MODELOS DE IA DE USO GENERAL 794
   2.1. Definiciones legales y clasificación 794
   2.2. La exigencia general de transparencia y una regulación singular para los modelos de GPAI 796
   2.3. Marco regulatorio propio 798
3. EL RIESGO EN LOS MODELOS Y SISTEMAS GPAI ¿CRITERIO SUFICIENTE PARA FIJAR LA OBJETIVACIÓN DE LA RC? 807
   3.1. Definiciones sobre el riesgo. Identificar incidente y peligro de IA 810
   3.2. ¿A qué sujetos se dirigen las obligaciones de evitar el riesgo? ¿A qué herramientas? 811
4. REFLEXIONES FINALES 814
5. BIBLIOGRAFÍA 816

Responsabilidad por conductas discriminatorias derivadas de los sesgos en el uso de la inteligencia artificial: jurisprudencia y reglamento europeo 817

Alberto Muñoz Villarreal

1. INTRODUCCIÓN 817
2. ANÁLISIS JURISPRUDENCIAL 818
3. EL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL 829

BIBLIOGRAFÍA 834

Inteligencia artificial y responsabilidad civil: un enfoque ético en la era digital 837

Iñigo A. Navarro MENDIZÁBAL

1. INTRODUCCIÓN 837
2. PRINCIPIOS ÉTICOS DE LA IA 840
   - **2.1. La importancia de la Ética en la IA** 840
   - **2.2. Principales principios éticos** 847
3. INTENTO DE APORTAR SOLUCIONES A LOS DESAFÍOS A LOS QUE SE ENFRENTA LA RC POR DAÑOS CAUSADOS POR LA IA 859
   - **3.1. RC objetiva o subjetiva** 859
   - **3.2. La Explicabilidad y Opacidad de los Sistemas de IA (Black Box)** 862
   - **3.3. Difusión de la Responsabilidad** 866
   - **3.4. Autonomía de la IA y Responsabilidad Humana** 869
   - **3.5. Daños colectivos y difusos** 871
   - **3.6. Daños futuros e inciertos** 873
4. BIBLIOGRAFÍA UTILIZADA 874

Los sistemas de inteligencia artificial, ¿productos defectuosos? 879

Manuel Ortiz Fernández

1. CUESTIONES PRELIMINARES 879
2. LA LEY DE INTELIGENCIA ARTIFICIAL 885
   - **2.1. Concepto y características básicas de la inteligencia artificial** 885
   - **2.2. El riesgo y la intervención humana: las actividades prohibidas y la clasificación de los sistemas** 893
3. LA RESPONSABILIDAD CIVIL DERIVADA DEL USO DE SISTEMAS INTELIGENTES 898
   - **3.1. Las relaciones entre las dos propuestas de Directiva** 898
   - **3.2. La responsabilidad civil en la (revisada) propuesta de Directiva sobre productos defectuosos** 903
   - **3.3. La propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial y las presunciones** 914

BIBLIOGRAFÍA 918

Perspectiva y categorización del riesgo en el Reglamento de Inteligencia Artificial 923

Miquel Peguera

1. INTRODUCCIÓN 923
2. LA PERSPECTIVA DEL RIESGO 926
3. LA PROHIBICIÓN DE PRÁCTICAS DE IA QUE IMPLICAN UN RIESGO EXCESIVO 930
4. SISTEMAS DE IA DE ALTO RIESGO VINCULADOS A LA LEGISLACIÓN ARMONIZADA SOBRE SEGURIDAD DE PRODUCTOS 935
5. SISTEMAS DE IA DE ALTO RIESGO INDEPENDIENTES 937
   - **5.1. Ejemplos de casos de uso relevantes** 939
   - **5.2. Criterios para rechazar la calificación de riesgo alto** 941
   - **5.3. Modificaciones de la relación de casos del Anexo III** 944
6. OBLIGACIONES DE TRANSPARENCIA FRENTE A RIESGOS DE CONFUSIÓN 944
7. RIESGOS SISTÉMICOS DE LOS MODELOS DE USO GENERAL 946

Inteligencia artificial generativa y daños por infracciones normativas del derecho de protección de datos personales. Un análisis a partir de la jurisprudencia reciente del TJUE sobre el artículo 82 RGPD 949

Antoni Rubí Puig

1. INTRODUCCIÓN 950
2. FUNCIONAMIENTO DE LA IA GENERATIVA E IMPLICACIONES PARA EL DERECHO DE PROTECCIÓN DE DATOS PERSONALES 954
   - **2.1. Concepto** 954
   - **2.2. Tipología** 955
   - **2.3. Cadena de valor** 956
3. CUESTIONES Y PROBLEMAS SOBRE LA REPARACIÓN DE DE DAÑOS 968
   - **3.1. Introducción: el artículo 82 RGPD como fundamento de responsabilidad civil** 968
   - **3.2. Daños mínimos y de bagatela** 970
   - **3.3. Indemnizabilidad del temor** 972
   - **3.4. Brechas de seguridad** 977
   - **3.5. Relaciones con otros fundamentos de responsabilidad: el caso de los *deepfakes*** 980
   - **3.6. Pluralidad de sujetos responsables** 983

4. CONCLUSIONES 985
BIBLIOGRAFÍA UTILIZADA 986
JURISPRUDENCIA DEL TJUE 990

El seguro de responsabilidad civil profesional de los operadores de sistemas de inteligencia artificial 993

Alberto J. Tapia Hermida

1. INTRODUCCIÓN 994
2. ANTECEDENTES 995
   2.1. La Resolución del Parlamento Europeo sobre un régimen de responsabilidad civil en materia de inteligencia artificial de 20 de octubre de 2020 995
   2.2. La Propuesta de Directiva sobre responsabilidad en materia de inteligencia artificial de 28 de septiembre de 2022 997
3. EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL 998
4. LAS CARACTERÍSTICAS DEL SEGURO DE RESPONSABILIDAD CIVIL DE LOS OPERADORES DE SISTEMAS DE INTELIGENCIA ARTIFICIAL 999
   4.1. Seguro voluntario 999
   4.2. Seguro de responsabilidad civil empresarial o profesional 1000
5. LAS PARTES 1000
   5.1. El asegurador 1000
   5.2. El tomador y el asegurado. Las pólizas colectivas 1001
6. EL RÉGIMEN DEL SEGURO DE RESPONSABILIDAD CIVL DE LOS OPERADORES DE SISTEMAS DE INTELIGENCIA ARTIFICIAL 1001
   6.1. Seguro de régimen común o seguro por grandes riesgos 1001
   6.2. Aplicación de la LCS 1002
   6.3. Aplicación de la LOSSEAR 1002
7. LA DELIMITACIÓN SUSTANCIAL DEL RIESGO CUBIERTO POR REFERENCIA A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL 1003
   7.1. Definición general del riesgo cubierto 1003
   7.2. Descripción específica de los riesgos excluidos de la cobertura 1003
8. LA DELIMITACIÓN TEMPORAL DEL RIESGO CUBIERTO POR REFERENCIA A LAS RECLAMACIONES PRESENTADAS CONTRA EL OPERADOR DE SISTEMAS DE INTELIGENCIA ARTIFICIAL ASEGURADO. LAS CLÁUSULAS "CLAIMS MADE" 1004

9. LA DEFENSA JURÍDICA DEL OPERADOR DE SISTEMAS DE INTELIGENCIA ARTIFICIAL ASEGURADO FRENTE A LA RECLAMACIÓN DEL USUARIO PERJUDICADO O DE SUS HEREDEROS 1006
10. LA ACCIÓN DIRECTA DEL USUARIO DE UN SISTEMA DE INTELIGENCIA ARTIFICIAL PERJUDICADO O SUS HEREDEROS CONTRA EL ASEGURADOR DEL OPERADOR 1007
11. LA TRANSPARENCIA DE LAS CONDICIONES DEL SEGURO DE RESPONSABILIDAD CIVIL DE LOS OPERADORES DE SISTEMAS DE INTELIGENCIA ARTIFICIAL..... 1008
12. CONCLUSIONES..... 1008

Transparencia y explicabilidad para prevenir la discriminación de los sistemas de inteligencia artificial: la interacción entre el RGPD y el RIA[1]

Mª Cristina Berenguer Albaladejo

Profesora Titular de Derecho civil
Universidad de Alicante

Sumario: 1.- LA DISCRIMINACIÓN ALGORÍTMICA COMO UNO DE LOS PRINCIPALES RIESGOS DERIVADOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL PARA LA TOMA DE DECISIONES; 2.- LA OPACIDAD COMO PRINCIPAL ESCOLLO PARA DETECTAR Y DEMOSTRAR LA DISCRIMINACIÓN ALGORÍTMICA; **2.1 Consideraciones previas; 2.2 Opacidad en el uso y sobre el contenido de los algoritmos; 2.3. Opacidad jurídica y técnica del algoritmo;** 3.- TRANSPARENCIA ALGORÍTMICA Y EXPLICABILIDAD: ¿QUÉ IMPLICAN ESTAS EXIGENCIAS?; 4.- MEDIDAS PARA GARANTIZAR LA TRANSPARENCIA Y LA EXPLICABILIDAD EN LA TOMA DE DECISIONES ALGORÍTMICAS; **4.1 Estado de la cuestión; 4.2. La transparencia y la explicabilidad en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, de protección de datos (RGPD): especial referencia a las decisiones automatizadas del art. 22;** *4.2.1. Ámbito de aplicación del art. 22 RGPD; A) Excepciones; B) La prohibición general del art. 22 RGPD: presupuestos de aplicación a la luz de la última jurisprudencia del TJUE en su sentencia de 7 de diciembre de 2023; 4.2.2. Contenido del derecho de información de los arts. 13.2 g), 14.2 g) y 15.1 h);* **4.3. La transparencia y la explicabilidad en el Reglamento**

[1] Trabajo realizado al amparo del Proyecto de Investigación «La irrupción de la inteligencia artificial en el Derecho de daños y su adaptación a las nuevas tecnologías», referencia PID2020-116185GB-100 (Ministerio de Ciencia e Innovación - Agencia Estatal de investigación) y del Proyecto «La nueva era de los algoritmos y la inteligencia artificial y su tutela jurídico-privada en el marco de la Unión Europea», referencia CIPROM/2022/40 (Programa Prometeo para Grupos de Investigación de Excelencia 2023 de la Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital).

(UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de IA; *4.3.1 Obligaciones de transparencia y comunicación de información*; A) Sistemas de alto riesgo; B) Determinados sistemas de IA con independencia de su riesgo; *4.3.2 Derecho a una explicación*; 5. CONSIDERACIONES FINALES SOBRE LA NECESIDAD DE TRANSPARENCIA Y EXPLICABILIDAD PARA DETECTAR Y DEMOSTRAR LA DISCRIMINACIÓN ALGORÍTMICA.

1. LA DISCRIMINACIÓN ALGORÍTMICA COMO UNO DE LOS PRINCIPALES RIESGOS DERIVADOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL PARA LA TOMA DE DECISIONES

Se ha definido la discriminación algorítmica como aquella que se produce cuando un individuo o grupo recibe un tratamiento arbitrario o injusto como consecuencia de la toma de decisiones automatizadas[2]. La principal diferencia con la discriminación que podríamos llamar «analógica» es la forma de llevarla a cabo, en tanto que la primera implica el uso de algoritmos y también, usualmente, de tecnología inteligente. No obstante, la conducta discriminatoria sería idéntica en ambos casos siempre que la decisión adoptada (bien por el humano, bien por el sistema de inteligencia artificial -en adelan-

[2] Así, MERCADER UGUINA, J.R. «Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del big data», *Labos: Revista de Derecho del Trabajo y Protección Social*, vol. 2, n. 2, 2021, p. 6. Con más detalle la define CASTILLO OCARANZA, C., «Discriminación algorítmica en el ámbito laboral», en P. Rivas Vallejo (dir.), *Discriminación algorítmica en el ámbito laboral*, Navarra 2022, p. 72, para quien existe discriminación algorítmica cuando un «algoritmo o sistema computacional algorítmico (X) produce discriminación grupal para una persona (Y) respecto de otra (Z), es decir, produce un resultado peor para la persona (Y) respecto a la persona (Z) debido a que (Y) pertenece a un cierto grupo social saliente o «socially salient group» (entendiendo por tal, aquél que pueda ser fácilmente identificado y que sea tal que pertenecer o no a este grupo tenga consecuencias para las relaciones sociales, y en particular, las relaciones de poder -ej. grupos delimitados por género, etnia, raza, discapacidad etc). Además, (X) debe haber basado su decisión en una observación estadísticamente relevante, es decir, en cualquier información que (X) utilice basándose en los datos de entrada (ej. si un empleador no contrata a una persona por tener la característica de ser mujer -que es un grupo social saliente- y contrata a otra que no es mujer basándose en la observación estadística de que es más probable que una mujer pida una baja por maternidad». Otras definiciones de este fenómenos se recogen en: FERNÁNDEZ DE LA MORENA, B., *Discriminación algorítmica. Estudio del sesgo en arquitecturas de aprendizaje profundo*, UAM, Madrid, 2019, p.1; CASTELLANOS CLARAMUNT, J., «Derecho e inteligencia artificial: atención especial a los sesgos, la privacidad y la protección de datos», accesible en https://idpbarcelona.net/derecho-e-inteligencia-artificial-atencion-especial-a-los-sesgos-la-privacidad-y-la-proteccion-de-datos/, la define como «un fenómeno que se produce cuando sistemas de IA o algoritmos utilizados para tomar decisiones automatizadas discriminan de manera indirecta o inconsciente a ciertas personas o grupos»; o GINÈS i FABRELLAS, A., «Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica», en RIVAS VALLEJO, P. (dir), *Discriminación algorítmica en el ámbito laboral*, Navarra, 2022, p. 302.

te, IA-) diera lugar a un trato diferente e injusto en situaciones comparables, o a un trato idéntico en situaciones distintas, basado en motivos protegidos y sin que hubiese justificación objetiva y razonable para ello[3].

Se ha destacado reiteradamente por las instituciones y la doctrina que uno de los mayores riesgos del uso de los sistemas de IA es la adopción de decisiones sesgadas y discriminatorias que, además de afectar a las personas concretas objeto de la decisión, lleven a reproducir y perpetuar estereotipos y desigualdades contra los que se lleva siglos luchando. Desde la *Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)),* hasta el reciente *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial* (en adelante, RIA)[4], ha sido constante el reconocimiento de que el derecho fundamental a la igualdad de trato y la no discriminación, junto con el de protección de datos personales y la intimidad, son especialmente vulnerables a través de esta tecnología y merecedores de una tutela especial, aumentado el riesgo de que sean vulnerados de forma proporcional al grado de autonomía del sistema utilizado, ya que el humano puede llegar a perder la dirección y control sobre las decisiones adoptadas, sin que además se pueda llegar a saber en base a qué motivos el sistema decide de una manera determinada[5]. De hecho, el peligro de que los sistemas de IA puedan perpetuar

---

[3] CONSEJO DE EUROPA, *Manual de legislación europea contra la discriminación* (edición 2018), Luxemburgo, 2019, p. 46 y ss. Los denominados «motivos» o «características protegidas» son aquellas condiciones de la persona que no se deben considerar relevantes para administrarle un trato diferenciado u otorgarle un determinado beneficio. Tanto es así, que no cualquier desigualdad de trato conlleva discriminación prohibida sino sólo la basada en tales características o motivos. Vid., ARCOS VARGAS, M., «La no discriminación en el Derecho derivado de la Unión Europea» en J.M. MORALES ORTEGA (dir.), *Realidad social y discriminación. Estudios sobre diversidad e inclusión laboral,* Murcia, 2022, pp. 26 y ss.

[4] DOUE de 12 de julio de 2024.

[5] Otros documentos que recogen esta idea son, entre otros muchos, el *Libro Blanco sobre Inteligencia artificial*, de 19 de febrero de 2020, p. 13, donde se mencionan los riesgos de la IA para los derechos fundamentales en general, haciendo especial referencia a la protección de los datos personales y a la no discriminación; la *Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL))*, que concibe la discriminación como una de las posibles causas en la producción de «lesión o daño» por estas tecnologías en el artículo 4 del Reglamento que la acompaña, estando presente en toda la propuesta el derecho a la no discriminación como uno de los pilares básicos del Derecho de la Unión; la *Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO* (2021), también presta especial atención a los retos que plantean las decisiones algorítmicas porque pueden «reproducir y reforzar los sesgos existentes, lo que puede exacerbar las formas ya existentes de discriminación, los prejuicios y los estereotipos». Y el GT29 en sus *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, 2018, subraya el peligro de que tanto la elaboración de perfiles como las decisiones automatizadas perpetúen los estereotipos existentes y la segregación social y lleven a discriminación injustificada (recuperado de

patrones históricos de discriminación contra las mujeres, ciertos grupos de edad, las personas con discapacidad o las personas de orígenes raciales o étnicos concretos o con una orientación sexual determinada, entre otros, es tenido en cuenta en el RIA para clasificar ciertos sistemas como de alto riesgo[6]. Tanto es así que en la Exposición de Motivos que contenía la Propuesta de RIA en su versión inicial, de 21 de abril de 2021, se declaraba expresamente que uno de los objetivos del establecimiento de requisitos específicos para los sistemas de alto riesgo era «*reducir al mínimo el riesgo de discriminación algorítmica, en particular en lo tocante al diseño y la calidad de los conjuntos de datos empleados para desarrollar sistemas de IA*», reconociendo con ello las principales causas que pueden provocarla[7].

Previamente al RIA, las *Directrices éticas para una IA fiable del Grupo Independiente de Expertos de Alto Nivel sobre IA* de 2019 (en adelante, Directrices éticas) ya situaban la «*necesidad de evitar el sesgo algorítmico injusto*»[8] entre los siete requisitos clave que debían cumplir los sistemas IA para ser considerados fiables. Tal como indica dicho documento, al que haremos especial referencia en este trabajo, la igualdad en el contexto de la IA implica la necesidad de evitar que el funcionamiento de este tipo de sistemas genere resultados injustamente sesgados, para lo que será preciso que los datos utilizados en su aprendizaje sean lo más inclusivos posibles de forma que estén representados los diferentes grupos de población[9].

Asimismo, la doctrina ha identificado los ataques a la privacidad y la discriminación como amenazas especialmente significativas que derivan del uso de sistemas de IA para la toma de decisiones[10]. Reflejo de lo apuntado es que

---

https://www.aepd.es/sites/default/files/2019-12/wp251rev01-es.pdf). Destacan esta idea también la *Carta de Derechos Digitales española* (2021) o la *Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital* (2023/C 23/01).

[6] Vid., por ejemplo, los Considerandos 56 y 57. También en el Considerando 67 se reconoce expresamente que los sistemas de IA *«perpetúan y amplifican la discriminación existente, en particular con respecto a las personas vulnerables pertenecientes a determinados grupos, en particular grupos raciales o étnicos»*.

[7] Vid., ap.1.2.

[8] Aunque «sesgo algorítmico» y «discriminación algorítmica» no son conceptos idénticos o coincidentes, en la práctica generalizada se emplean indistintamente para hacer referencia a las desviaciones de los sistemas de IA que conllevan resultados discriminatorios en forma de predicciones, recomendaciones o decisiones. Precisando las diferencias entre ambos conceptos, GERARDS, J./ XENIDIS, *op.cit.*, p. 49.

[9] Pp. 13 y 23. Disponible en: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

[10] Así se pone de manifiesto, por ejemplo, en EUBANKS, V., *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor,* St MARTÍN›s Press, 2018; O'NEIL, C., *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* Capitán Swing, Madrid, 2017; GERARDS, J./ XENIDIS, R., *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law* (Special report European network of legal experts in gender equality and non-discrimination), European Commission, 2020; MANHEIM, K. M., & KAPLAN,

cada vez contamos con más evidencias de que las decisiones automatizadas no son siempre neutrales ni objetivas, sino que, muy al contrario, el riesgo de llegar a resultados parciales, sesgados o discriminatorios es más elevado de lo que, *a priori*, pudiera parecer. Algunos de los casos más mediáticos de discriminación algorítmica acaecieron hace ya algunos años en el seno de grandes empresas como Apple, Amazon o IBM, cuyos *softwares* demostraron tener prejuicios sexistas o racistas al haber sido entrenados con datos sesgados y tuvieron que ser retirados. Respecto de otros que también tuvieron una gran repercusión, contamos incluso con pronunciamientos judiciales que vienen marcando el rumbo a seguir en la materia. Así, el caso del algoritmo *Syri (Systeem Risico Indicatie*, o Sistema Indicativo del Riesgo) resuelto por el Tribunal de lo Civil de la Haya en su sentencia de 5 de febrero de 2020; el caso del algoritmo *Frank*, resuelto por el Tribunal Ordinario de Bolonia en sentencia de 31 de diciembre de 2020; o el caso *Loomis v. Wisconsin*, conocido por el Tribunal Supremo de Wisconsin (EEUU), que giró en torno a un sistema inteligente de evaluación de riesgos denominado *COMPAS* (*Correctional Offender Management Profiling for Alternative Sanctions*). En todos ellos se evidenció que el algoritmo (o el uso que se hacía del sistema) discriminaba a ciertos colectivos[11].

L., «Artificial Intelligence: Risks to Privacy and Democracy», Yale Journal of Law and Technology 106, vol. 21, 2019, pp. 106-189, disponible en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3273016; GINÈS i FABRELLAS, A., «Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica», en RIVAS VALLEJO, P. (dir), *Discriminación algorítmica en el ámbito laboral*, 2022; o los trabajos de SORIANO ARNANZ, A. «Decisiones automatizadas y discriminación: aproximación y propuestas generales», *Revista General de Derecho Administrativo*, Nº. 56, 2021, «Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos», *Revista de Derecho Público: teoría y método*, Nº13, 2021, pp. 85-127, y «La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de Inteligencia Artificial», en *Nuevas normatividades: inteligencia artificial, derecho y género* (coord. por P.R. Bonorino Ramírez / R. Fernández Acevedo / P. Valcárcel Fernández), Navarra, 2022, pp. 63-88; MUÑOZ GUTIÉRREZ, C., «La discriminación en una sociedad automatizada: Contribuciones desde América Latina», *Revista chilena de derecho y tecnología*, 10(1), 2021, p. 273; SÁNCHEZ CAPARRÓS, M. «Prevenir y controlar la discriminación algorítmica», 2022. Recuperado de https://www.researchgate.net/publication/358207305_Prevenir_y_controlar_la_discriminacion_algoritmica, p. 17/34; o GASCÓN MARCÉN, A., «Derechos humanos e inteligencia artificial», en *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Pérez Miras (dir.), Teruel Lozano (dir.), Raffiotta (dir.), Pia Iadicco (dir.), Vol. 5, 2020 (Retos en el siglo XXI / coord. por Silvia Romboli), p. 337.

[11] Otros ejemplos que demuestran que en la práctica estos casos son más frecuentes de lo que sería deseable, los protagonizaron Google o Microsoft. En el primer caso, cuando el algoritmo de clasificación de *Google Photos* después de su aprendizaje automático confundió a varios jóvenes afroamericanos con gorilas o cuando se demostró que su algoritmo publicitario mostraba ofertas de empleo mejor remuneradas a los hombres que a las mujeres. En caso de Microsoft, cuando el *chatbot Tay* -que se servía de la IA para mantener conversaciones en una conocida red social con un público de entre 18 y 24 años-, empezó a elaborar mensajes de contenido racista, sexista y xenófobo, y tuvo que ser retirado por la compañía al día siguiente de su lanzamiento. También en el sector de la conducción autónoma se ha observado el sesgo racial, al demostrar un estudio llevado a cabo por investigadores del Instituto Tecnológico de Georgia (EEUU) sobre ocho sistemas de inteligencia artificial emplea-

Sin duda, el empleo de sistemas de IA para la toma de decisiones ha hecho que la discriminación adquiera unas dimensiones y afronte unos retos hasta el momento inexistentes[12]. De hecho, los expertos vienen advirtiendo desde hace años sobre la potencialidad de la tecnología para *extraer y procesar datos,* pero también para *reproducir y perpetuar discriminaciones existentes*, *amplificar sesgos* y, lo que es aún más peligroso, para *hacer que la discriminación pase desapercibida*[13]. De ahí que se haya mantenido, con acierto a nuestro juicio, que uno de los rasgos definitorios y más problemáticos de la discriminación algorítmica es precisamente su *invisibilidad*, junto con su intensidad y su complejidad técnica[14]. Dichas características y sus consiguientes desafíos, justifican que se haya convertido en un fenómeno nuevo y distinto que obliga a hacer ajustes en los mecanismos de tutela antidiscriminatoria existentes incorporando exigencias de transparencia y explicabilidad desde el diseño, evaluaciones de impacto que midan el potencial de discriminación del sistema por diversas causas (género, etnia, religión etc.) y auditorías, con el fin de poder detectar

---

dos en la detección de objetos, que tendían a favorecer los tonos claros de piel sobre los oscuros y discriminaban a los peatones en función de dicho rasgo. Y en otro estudio realizado respecto de un algoritmo utilizado por aseguradoras y hospitales en EEUU para analizar los riesgos de salud de los pacientes y asignar atención médica, también se advirtió que el sistema discriminaba sistemáticamente a la población afroamericana.

[12] En este sentido, por todos, GINÈS i FABRELLAS, A., *op.cit.,* pp. 316 y ss. También GERARDS, J./ XENIDIS, R., *op.cit.,* p. 75, ponen de manifiesto que, si bien muchos de los problemas que plantea la discriminación existían antes del desarrollo de los algoritmos y han sido identificados por la doctrina, el uso creciente de algoritmos en todos los ámbitos de la sociedad los agudiza y multiplica.

[13] Estos peligros han sido puestos de manifiesto por la mayoría de la doctrina que ha tratado el tema. Pueden verse, entre otros, los trabajos citados *supra* en nota 10. Con detalle explica cada uno de ellos, GINÈS i FABRELLAS, A., *op.cit.,* pp. 315-321. También ALAMEDA CASTILLO, M.T., «Reclutamiento tecnológico. Sobre algoritmos y acceso al empleo», *Temas laborales: Revista andaluza de trabajo y bienestar social*, Nº 159, 2021, p. 14. Ya en el año 2017, el Parlamento Europeo alertaba en su *Resolución de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* (2015/2103(INL)) de que el aprendizaje automático ofrecía enormes ventajas económicas a la sociedad al mejorar enormemente la capacidad de analizar datos, pero también planteaba retos a la hora de velar por la no discriminación, las garantías procesales, la transparencia y la inteligibilidad de los procesos decisorios. Vid., Considerando letra h). También en el apartado sobre Principios éticos señala que «el potencial de empoderamiento que encierra el recurso a la robótica se ve matizado por una serie de tensiones o posibles riesgos y que debe ser evaluado detenidamente a la luz de la seguridad y la salud humanas; la libertad, la intimidad, la integridad y la dignidad; la autodeterminación y la no discriminación, y la protección de los datos personales» (nº 10).

[14] Así lo mantienen, entre otros, SÁEZ LARA, C., «El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación», *Temas laborales: Revista andaluza de trabajo y bienestar social*, Nº 155, 2020, p. 45, para quien este rasgo es, a su vez, el principal problema de estas tecnologías; ALAMEDA CASTILLO, M.T., *op.cit.* p. 14, siguiendo a la primera; o EGUÍLUZ CASTAÑEIRA. J.A., «Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales», *Estudios de Deusto: revista de Derecho Público*, vol. 68, n. 2, 2020, p. 337.

la discriminación, prevenirla o corregirla y, en última instancia, reparar los daños que ocasione[15].

Precisamente a las medidas dirigidas a garantizar la transparencia y explicabilidad de los sistemas de IA para detectar y prevenir errores, sesgos y discriminación injusta dedicaremos el presente trabajo, que comenzará abordando la opacidad algorítmica en sus diversas manifestaciones como principal obstáculo para la consecución de dichos objetivos. Tras ello, se analizará qué significa concretamente que los sistemas de IA deban ser transparentes y explicables y qué soluciones se han venido adoptando para conseguirlo, dedicando particular atención a las previstas en el RGPD y en el RIA.

2. LA OPACIDAD COMO PRINCIPAL ESCOLLO PARA DETECTAR Y DEMOSTRAR LA DISCRIMINACIÓN ALGORÍTMICA

2.1. Consideraciones previas

Los principales desafíos que plantean los sistemas de IA tanto para la salud, la seguridad y los derechos fundamentales en general, como para la igualdad y no discriminación en particular, están directamente relacionados con sus características intrínsecas. Su complejidad técnica, interconectividad, opacidad, vulnerabilidad, apertura, autonomía y capacidad de autoaprendizaje, así como la pluralidad de agentes involucrados en su diseño, desarrollo, implementación y uso, dificultan la averiguación de *cuándo, cómo o por qué* se ha originado el error del sistema y *quién debe responder* de los daños que se ocasionen[16].

---

[15] Para MERCADER UGUINA, J.R., *op.cit.*, p. 6, aunque por el momento la discriminación algorítmica obtiene respuestas desde nuestras actuales técnicas conceptuales, requiere que se adopten nuevas herramientas jurídicas de análisis.

[16] El Grupo de Expertos en responsabilidad civil y nuevas tecnologías (*Expert Group on Liability and New Technologies - New Technologies Formation*, conocido por sus siglas en inglés NTF), en su Informe titulado *Liability for artificial intelligence and other emerging digital technologies*, de 21 de noviembre de 2019, explicaba una serie de características comunes a las tecnologías digitales emergentes en general que podían complicar la determinación de la responsabilidad civil cuando se ocasionaran daños. Dichas características eran las siguientes: (a) complejidad, (b) opacidad, (c) apertura, (d) autonomía, (e) previsibilidad, (f) manejo y dependencia de datos, y (g) vulnerabilidad. Estas características también se recogen por la Comisión Europea en su *Informe al Parlamento Europeo, al Consejo y al Comité Económico y Social Europeo, sobre las repercusiones en materia de seguridad y responsabilidad civil de la inteligencia artificial, el internet de las cosas y la robótica, de 19 de febrero de 2020* (COM/2020/64 final). Explica cada uno de estos rasgos, por todos, ATIENZA NAVARRO, M.L, *Daños causados por inteligencia artificial y responsabilidad civil*, 2022, pp. 56-66.

A nuestro modo de ver, de todas estas características es la *opacidad* la que se sitúa en el origen de la mayoría de los problemas mencionados, ya que, la falta de transparencia tanto a la hora de *utilizar* estos sistemas para tomar decisiones que afectan a personas concretas, como a la hora de *explicar los motivos* que las sustentan, hace que sea extraordinariamente difícil para el ciudadano averiguar y demostrar que está siendo discriminado[17].

Ya advertía el Parlamento Europeo en su *Resolución de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial* (2020/2014(INL))[18], que la opacidad algorítmica podía hacer extremadamente costoso, o incluso imposible, determinar *quién controlaba el riesgo asociado al sistema de IA,* o *qué código, entrada o datos provocaban en última instancia el funcionamiento lesivo,* o *la identificación de la relación de causalidad entre el daño y el comportamiento que lo causa*, con el resultado de que las víctimas podían quedar sin recibir una indemnización adecuada. Y la misma idea la reitera la Comisión a lo largo de su última *Propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial,* de 28 de septiembre de 2022 (COM/2022/496 final)[19].

Para Lazcoz Moratinos la opacidad es la «preocupación que más atención ha despertado en la literatura al evaluar la aplicación de estos modelos algorítmicos en la toma de decisiones»[20]. Es por ello que, como se irá viendo a lo largo de estas páginas, en todos los instrumentos normativos vigentes o proyectados relacionados con la IA, tanto en el ámbito europeo como fuera de él, la transparencia se erige en pieza clave o principio rector a la hora de diseñar, desarrollar e implementar algoritmos para la toma de decisiones. De

---

[17] En el estudio realizado por GERARDS, J./ XENIDIS, R., *op.cit.,* p. 108, se destaca que los problemas de discriminación de los sistemas de IA pueden ser difíciles de detectar y confrontar debido principalmente a problemas de transparencia y responsabilidad.

[18] Vid., Considerando H).

[19] Vid., por ejemplo, pp. 1, 16 y Considerandos 3 o 27. En el Considerando 3 señala que «cuando la IA se interpone entre el acto u omisión de una persona y el daño, las características específicas de determinados sistemas de IA, como la opacidad, el comportamiento autónomo y la complejidad, pueden hacer excesivamente difícil, si no imposible, que el perjudicado satisfaga la carga de la prueba. En particular, puede resultar excesivamente difícil demostrar que un dato de entrada concreto del que es responsable la persona potencialmente responsable ha dado lugar a una información de salida específica de un sistema de IA que, a su vez, ha provocado el daño en cuestión».

[20] Así lo afirma LAZCOZ MORATINOS, G., «Análisis jurídico de la toma de decisiones algorítmica en la asistencia sanitaria», en *La regulación de los algoritmos*, coord. por G.M. Díaz González/ A.J. Huergo Lora (dir.), 2020, p. 286. También TOLOSA, P./DIBO, C., «Inteligencia artificial, discriminación por género y Derecho: viejos problemas, nuevos desafíos», en C. Danesi (dir.) *Inteligencia Artificial, tecnologías emergentes y Derecho,*2021, p.187, consideran que los algoritmos deben ser auditables, transparentes, y explicables y si se logra implementar algoritmos con tales características seguramente las decisiones que se obtengan podrán ser más transparentes que las decisiones humanas bien intencionadas, pero inconscientemente sesgadas.

hecho, en las Directrices éticas del Grupo de Expertos en IA se incluye como uno de los siete requisitos esenciales a cumplir por los sistemas para conseguir una IA fiable. Y en la *Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)* -en adelante, Resolución PE sobre aspectos éticos de la IA 2020-, se estableció que el marco regulador de la IA debía regirse por los principios de transparencia, explicabilidad, equidad, rendición de cuentas y responsabilidad, con el fin de potenciar las ventajas de los sistemas y proteger al mismo tiempo a los ciudadanos de los riesgos que implican. Dicha protección, tal como subraya el Parlamento, implica que los humanos puedan comprender las acciones de la IA, para lo cual es imprescindible que los desarrolladores e implementadores garanticen, dentro de los límites de lo técnicamente posible, que la tecnología se despliega y utiliza respetando los requisitos de transparencia y permitiendo la auditoría y la trazabilidad.

2.2. Opacidad en el uso y sobre el contenido de los algoritmos

Conviene precisar, de entrada, que la opacidad algorítmica puede y suele manifestarse tanto a la hora de revelar que se están usando estos sistemas para decidir cuestiones concretas, como a la hora de proporcionar información sobre ellos que permita a los interesados saber cómo y en base a qué motivos se ha adoptado la decisión que les afecta. Es decir, en muchas ocasiones las personas ignoramos que estamos siendo objeto de decisiones algorítmicas y de que, por ejemplo, las mismas nos discriminan injustamente. Así, una persona que ha sido discriminada por una plataforma de Internet en relación con el precio de determinados productos en función de su perfil (incluido su género, edad, etc.) puede que ni siquiera sepa que la oferta que se le muestra ha sido decidida por un sistema inteligente y menos aún que el mismo puede estar vulnerando sus derechos, sobre todo cuando no tiene una referencia comparativa, es decir, si no sabe que a otras personas que cuentan con atributos diferentes se le ofrecen los mismos productos a un mejor precio[21].

El Parlamento Europeo, en su *Resolución sobre aspectos éticos de la IA 2020,* hacía una sutil referencia a esta idea al distinguir entre la *transparencia de*

[21] Expone esta situación SORIANO ARNANZ, A. «Decisiones automatizadas y discriminación: aproximación y propuestas generales», *Revista General de Derecho Administrativo,* Nº. 56, 2021, p. 17/31, para quien, el acceso a la justicia para defender el derecho a la igualdad de trato y no discriminación no sólo se ve condicionado por cuestiones económicas o burocráticas, sino también por el hecho de que las víctimas de las situaciones de discriminación con frecuencia no son conscientes de que sus derechos están siendo vulnerados.

los algoritmos y la transparencia en el *uso de los algoritmos*[22]. Y también en su *Resolución de 12 de febrero de 2020, sobre los procesos automatizados de toma de decisiones: garantizar la protección de los consumidores y la libre circulación de bienes y servicios (2019/2915 (RSP)),* advertía de que el uso de procesos automatizados de toma de decisiones planteaba retos para los consumidores que podían afectar tanto a su capacidad para *detectar* dichos procesos, como de *entender* su funcionamiento, *tomar decisiones con conocimiento de causa* en cuanto a su utilización y *autoexcluirse*[23].

En virtud de lo anterior, consideramos junto a un sector de nuestra doctrina que habría que distinguir dos niveles de opacidad algorítmica[24]. El primero sería aquel en que el interesado desconoce que es objeto de una decisión automatizada, y el segundo, aquel en que sabe que la decisión que le afecta ha sido o va a ser adoptada por un sistema de IA pero no se le permite conocer el funcionamiento del proceso decisorio ni los motivos de la decisión que le afecta negativamente.

Como es lógico, si ni siquiera sabemos que estamos siendo objeto de decisiones algorítmicas, menos aún seremos capaces de descubrir que las mismas son discriminatorias y, en consecuencia, perderemos la oportunidad de impugnarlas en defensa de nuestros derechos[25]. De lo que no cabe duda es de que, en tales casos, la frustración de las legítimas expectativas de las personas o la pérdida de oportunidad de acceder a un empleo, conseguir un crédito, ingresar en un prestigioso centro educativo o ser beneficiario de un tratamiento médico o de un seguro de vida (entre otros muchos ejemplos que podrían ponerse) basadas en criterios injustos o discriminatorios, generan daños a personas concretas que en muchos casos quedarán sin resarcir por desconocimiento o ignorancia acerca de que, tras esas decisiones, se esconde un proceso algorítmico y de que el mismo arroja resultados discriminatorios.

---

[22] Vid., ap. 20.

[23] Vid., ap. C).

[24] LAZCOZ MORATINOS, G., *op.cit.,* p. 293, VESTRI, G., «La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativ», *Revista Aragonesa de Administración Pública*, nº 56, 2021, p. 382.

[25] En la misma línea considera LLORENS ESPADA, J., «Responsabilidades civiles por discriminación por razón de género cuando medie un sistema de inteligencia artificial», en Rivas Vallejo (dir.), *Discriminación algorítmica en el ámbito laboral*, Navarra 2022, p. 576, que en el ámbito laboral será difícil para un trabajador saber si una decisión empresarial esconde un proceso algorítmico, o incluso sabiéndolo, no será fácil obtener información de los motivos en base a los cuales el algoritmo toma las decisiones y de qué parte del proceso deriva la falta de equidad. Por ende, el trabajador no podrá saber que está siendo discriminado ni podrá incoar un procedimiento de tutela de sus derechos fundamentales ante casos de discriminación generados por algoritmos. En sentido parecido, Eguíluz Castañeira, J.U., *op.cit.,* p. 337.

Es más, aun siendo conscientes de que detrás de la decisión que nos afecta se esconde un proceso algorítmico, seguiría siendo muy difícil *detectar* posibles errores o casos específicos de discriminación teniendo en cuenta que la decisión algorítmica se basa en una combinación de múltiples variables con ponderaciones diversas a las que el ciudadano en principio no tiene acceso o, incluso teniéndolo, no comprende. Y es que en muchas ocasiones estos sistemas de IA son técnicamente tan complejos que ni siquiera los programadores pueden explicar las razones por las que ha llegado a una decisión y no a otra distinta.

Urge por tanto, como paso previo, que la ciudadanía tome plena conciencia de que hoy en día muchas de las decisiones que nos afectan -y que normalmente se basan en nuestros datos personales- están siendo *adoptadas por algoritmos o en base a ellos*, que pueden contener sesgos y arrojar resultados que vulneran nuestro derecho a la igualdad trato y a la no discriminación. Sólo así se podrá superar el primer escollo que presenta la discriminación algorítmica que no es otro que su detección o identificación.

2.3. Opacidad jurídica y técnica del algoritmo

Asimismo, en el tratamiento del problema de la opacidad algorítmica inciden factores tanto técnicos como jurídicos. Es decir, la falta de transparencia puede deberse tanto a la propia complejidad intrínseca del algoritmo, cuya comprensión exige una capacidad analítica y unos conocimientos técnicos considerables, como al régimen jurídico que lo protege en tanto en cuanto pueda ser objeto de derechos de propiedad intelectual o secreto empresarial[26]. Esta doble opacidad, que dificulta tanto el *acceso* como la *comprensión*

[26] La doctrina ha hecho referencia a esta idea de varias maneras. Por ejemplo, para LAZCOZ MORATINOS, G., *op.cit.*, p. 293, en el ámbito de la IA la opacidad es polisémica y puede utilizarse con un doble sentido: por un lado, como equivalente a *ininteligible* (inherente al *machine learning in a big data context*), y por otro lado, para designar la opacidad intencionada en base a razones de derechos de autor, secreto industrial, privacidad o seguridad. Para SUBERBIOLA GARBIZU, I., «Inteligencia artificial, opacidad y motivación de las decisiones administrativas automatizadas», en *La protección de los derechos fundamentales en el ámbito tributario*, coord. por A.Vázquez del Rey Villanueva e I. Suberbiola Garbizu, 2021, recuperado de La Ley Digital (LA LEY 1117/2021), hay que ser conscientes de que existen *black boxes* por la arquitectura del propio sistema de IA y *black boxes* inducidas por la normativa que lo regula. Y CASTILLO OCARANZA, C., «Discriminación algorítmica en el ámbito laboral», en P. Rivas Vallejo (dir.), *Discriminación algorítmica en el ámbito laboral*, Navarra 2022, p. 75, considera que la opacidad algorítmica podría ser incluso triple, ya que, junto a la opacidad legal y a la opacidad técnica debida a la complejidad inherente al propio sistema, habría que añadir la inducida a través de «mecanismos de ofuscación intencional» utilizados para alterar las instrucciones de un programa (esto es, su código fuente) de forma que, a pesar de seguir manteniendo su función original, se hicieran difíciles de leer.

del sistema, pone en tela juicio el derecho a obtener una resolución o decisión motivada y la posibilidad de recurrirla, generando indefensión para el afectado y un grave riesgo de arbitrariedad, sobre todo cuando estos sistemas son utilizados por la Administración puesto que en el ámbito público las exigencias de transparencia en los procesos decisorios se acrecientan.

Respecto a la opacidad técnica o consustancial al algoritmo, la misma aumenta cuando se emplean algoritmos de autoaprendizaje no supervisado y profundo que gozan de gran autonomía decisoria y cuya capacidad predictiva (altamente precisa por otro lado) reside en establecer correlaciones tan sumamente complejas que sus resultados pueden ser ininteligibles para el ser humano a partir de los datos que se aportan al modelo[27]. Este efecto, conocido como «caja negra» o *black box*, hace referencia a la imposibilidad de identificar las razones que han llevado al sistema de IA a adoptar sus decisiones. Por ello, el uso de esta tecnología plantea un problema estructural en la medida en que su esencia misma choca con las normas que exigen transparencia, trazabilidad y posibilidad de identificar los procesos de toma de decisiones de los algoritmos.

Respecto a este tipo de sistemas donde la lógica computacional es elusiva y opaca, ni siquiera la apertura de la caja negra y el acceso al código fuente del algoritmo garantizaría la comprensibilidad o inteligibilidad de los resultados[28]. Este problema estructural hace que la propia idea de la transparencia resulte cuestionable e incluso se hable de la «falacia de la transparencia»[29],

---

[27] Apunta BARRIOS ANDRÉS, M., *Manual de Derecho Digital*, Valencia, 2020, p. 62, que el término «profundo» hace referencia al gran número de capas de neuronas ocultas en la red. A diferencia del aprendizaje automático supervisado, en el que han de realizarse manualmente los procesos iniciales de extracción de características relevantes y diseño de un modelo capaz de categorizar o clasificar la información que se desea analizar, en los de aprendizaje profundo estos procesos de extracción de características y modelización son automáticos. Los hace el sistema a partir de imágenes, textos o sonidos. Son altamente precisos y en ocasiones superan el rendimiento humano.

[28] Así lo afirman, entre otros muchos, SOLAR CAYÓN, J.I., «Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos», en *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho,* J.I. Solar Cayón (ed.), Alcalá de Henares, 2020, p. 149; RIVAS VALLEJO, P. «Análisis desde el derecho antidiscriminatorio», en P. Rivas Vallejo (dir.*) Discriminación algorítmica en el ámbito laboral*, Navarra 2022, p. 441 EVANGELIO LLORCA, R., «Causalidad y responsabilidad civil por daños ocasionados por sistemas de inteligencia artificial: las presunciones de causalidad en las propuestas normativas de la UE», en *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*, coord. por N. Álvarez Lata; J.M. Busto Lago (pr.), 2024, p. 561. No obstante, como mantiene NAVAS NAVARRO, S., «Sistemas expertos basados en inteligencia artificial y responsabilidad civil. Algunas cuestiones controvertidas», *Diario La Ley,* 13 de Diciembre de 2019, LA LEY 14980/2019, p. 7/16, si bien el acceso a la *black box* no será la panacea para la víctima, quizá sí que sea de utilidad para fijar indicios.

[29] SOLAR CAYÓN, J.I., *op.cit.,* p. 148. También advierte el NIST (*National Institute of Standards and Technology*) que la transparencia no garantiza la explicabilidad, especialmente si el usuario no comprende los principios técnicos. Vid., *AI Risk Management Framework: Initial Draft*, 17 de marzo

ya que no siempre que se consigue acceder a la caja negra de los sistemas algorítmicos quedan salvaguardados adecuadamente los derechos de los interesados. Especialmente a los efectos de la asignación de responsabilidad, el acceso a la información de las cajas negras puede no ser suficiente para lograr una conexión entre el autor, el algoritmo y sus consecuencias, que sea lo suficientemente sustancial como para que alguien pueda considerarse responsable de sus efectos perjudiciales[30]. Pero es que, además, incluso cuando se pudiera detectar la causa del problema, la misma podría variar o desaparecer con el tiempo teniendo en cuenta que este tipo de algoritmo está en constante evolución[31].

Por estos motivos, un sector doctrinal considera que, aun a costa de sacrificar sus ventajas (principalmente, su mayor nivel de precisión), debe demorarse el empleo de este tipo de sistemas de aprendizaje profundo en los que no se puede garantizar el respeto de los derechos fundamentales[32]. De hecho,

---

de 2022. Accesible en https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

[30] GONZÁLEZ VALVERDE, A., «Responsabilidad por el potencial de sesgo discriminatorio de los algoritmos de los productos de Inteligencia Artificial. A propósito de la Algorithmic Accountability Act of 2019 (s.1108)». En Ataz López y Cobacho Gómez, *Cuestiones clásicas y actuales del Derecho de daños: Estudios en homenaje al profesor Dr. Roca Guillamón,* Vol. 2, 2021, pp. 1263-1264. Las instituciones europeas han venido advirtiendo de este *handicap* en diferentes documentos, destacando que dicho efecto de caja negra podría llegar a impedir que los afectados obtuvieran una indemnización por los daños causados por aplicaciones de IA autónomas. Vid., por ejemplo, el *Informe sobre las repercusiones en materia de seguridad y responsabilidad civil de la IA, el internet de las cosas y la robótica*, de 19 de febrero de 2020, de la Comisión.

[31] Precisamente por ello se venía abogando por la necesidad de que esta tecnología contara con un sistema que permitiera registrar información sobre su funcionamiento, esto es, datos relativos a su movimiento, actuación y decisión («logging by design» o registro desde el diseño) a modo de *caja negra* como la que se encuentra hoy en día en aviones, trenes o drones. Entre otros lo defendían, NAVAS NAVARRO, S., «La perspectiva de género en la inteligencia artificial», *Diario La Ley,* nº 48, Sección Ciberderecho, 8 de marzo de 2021, pp. 8 y 11/23, basándose en el Informe del *Expert Group on Liability and New Technologies, Liability*, o SORIANO ARNANZ, A., «La aplicación del marco jurídico europeo…», *cit.*, p. 81, para quien en los sistemas que se van actualizando, es necesario que los operadores guarden una copia de cada versión del programa utilizado en los procesos de toma de decisiones.

[32] En el ámbito europeo, el Informe *AI in the UK:ready, willing and able?*, elaborado por el Comité sobre inteligencia artificial de la Cámara de los Lores británica, recomienda que el empleo de sistemas de IA basados en redes neuronales profundas que puedan tener un impacto sustancial en la vida de los individuos se demore hasta que se encuentren soluciones técnicas que permitan su comprensibilidad plena (parágrafo 105). Accesible en: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm. En relación con esto, considera ALCOLEA AZCÁRRAGA, C., «La responsabilidad patrimonial de la Administración y el uso de algoritmos», *Revista General de Derecho Administrativo,* nº 59, enero 2022, p.8/31, la posibilidad de extrapolar al ámbito tecnológico el *principio de precaución* aplicable en Derecho ambiental. Con el fin de controlar las consecuencias imprevisibles de los sistemas de IA y evitar daños irreparables, esgrime que podría adoptarse el principio de precaución algorítmica, por cuanto el ámbito tecnológico tiene también un importante potencial dañino para la humanidad en su conjunto. En base a dicho principio de precaución concluye

la ininteligibilidad de algunos algoritmos de caja negra ha frenado su empleo en sectores como la medicina, limitando su uso como soporte o apoyo en la toma de decisiones clínicas[33]. No obstante, frente a esta cautelosa postura, la opinión mayoritaria aboga por la utilización de tales sistemas adoptando, eso sí, un enfoque de «explicabilidad» sobre la mera transparencia técnica, conceptos ambos que se analizarán el siguiente apartado[34]. Tanto es así, que existe actualmente una tendencia a desarrollar modelos algorítmicos más sencillos, transparentes y explicables que respondan mejor a las exigencias jurídicas y generen mayor confianza en los usuarios, que ha dado lugar a un campo de estudio e investigación denominado *Explainable Artificial Intelligence* o *Interpretable Machine Learning*, a través del que se pretende encontrar un equilibrio adecuado entre la optimización de la capacidad predictiva de los algoritmos y su explicabilidad[35]. De lo que se trataría, en definitiva, sería de integrar la transparencia como estrategia *ex ante* y la explicabilidad *ex post*, para que los usuarios de estos sistemas de IA pudieran comprender el porqué de sus decisiones y rebatirlas si lo desean.

Junto con esta opacidad estructural del algoritmo, nos encontramos con la opacidad jurídica en la medida en que el acceso al mismo (tanto a su código fuente como a los datos utilizados por el sistema de IA para funcionar y aprender) podría denegarse al interesado si se considera protegido por derechos de propiedad intelectual o como secreto empresarial[36].

---

el autor que, aunque algunos «autores consideran que la aplicación de los algoritmos a las facultades discrecionales de la Administración pública aportaría un baremo nuevo que alejaría la arbitrariedad, podría parecer que su uso sin una completa comprensión de cómo funcionan o por qué toman las decisiones que toman debería postergarse a un momento en que efectivamente se disponga de ese control y posibilidad de fiscalización». Recuperado de https://laadministracionaldia.inap.es/noticia.asp?id=1512629.

[33] LAZCOZ MORATINOS, G., *op.cit.*, p. 293.

[34] COTINO HUESO, L., «Transparencia de la inteligencia artificial pública…», *cit.*, señala, citando a Gutiérrez David, que la «transparencia técnica sería una característica pasiva del modelo de IA que permite al observador humano comprender o entender el sistema, mientras que habría otros sistemas de IA que no son transparentes, pero pueden llegar a ser explicables mediante distintas técnicas a partir del comportamiento del modelo, los datos utilizados, los resultados obtenidos y del proceso completo de la toma de decisión».

[35] Esta necesidad de encontrar un equilibrio entre la mejora de la explicabilidad de un sistema (que puede reducir su precisión) o una mayor precisión del mismo (a costa de la explicabilidad), ya se ponía de relieve en las *Directrices éticas* del Grupo de Expertos. En dicho documento se establecía la necesidad de que cuando un sistema de IA tuviera un impacto significativo en la vida de las personas, debería ser posible reclamar una explicación adecuada del proceso de toma de decisiones del sistema de IA (*vid.*, ap. 77).

[36] Se ha mantenido que los algoritmos podrían quedar protegidos por la *Ley 1/2019, de 20 de febrero de 2019, de secretos empresariales*, siempre que reunieran las condiciones necesarias para ser considerados como secreto empresarial, esto es, siempre que se tratase de información no divulgada por las empresas, que tuviera valor comercial y se hubieran adoptado medidas razonables para su protección. En este sentido, por ejemplo, la STS n. 508/2018, de 20 de septiembre, confirmó que

Asimismo, se han esgrimido razones de seguridad pública para no ofrecer información sobre los algoritmos utilizados por la Administración, impidiendo a los particulares afectados conocer el proceso decisorio y las razones por las que se ha llegado al resultado que les perjudica.

Así ocurrió, por ejemplo, en el polémico caso del algoritmo *Bosco* para la concesión del bono social eléctrico a personas vulnerables[37]. Ante las abundantes quejas de usuarios a los que se les había denegado la ayuda y creían tener derecho ella, la Fundación CIVIO solicitó a la Administración, al amparo de la Ley de Transparencia, que le proporcionara su especificación técnica, los resultados de las pruebas de comprobación de la aplicación y su código fuente, a fin de comprobar si el programa presentaba algún defecto o errores en su configuración que impedían la correcta aplicación de la legislación[38].

---

el código fuente de un programa informático podía estar protegido como secreto comercial si su divulgación suponía un perjuicio económico para su titular y siempre que se hubiera mantenido en secreto. Mayor polémica ha generado su protección por los derechos de autor regulados en el título VII LPI. Aunque no podemos profundizar ahora en este asunto, simplemente mencionaremos que hay quien defiende que la definición de «programa de ordenador» del art. 96.1 LPI es aplicable al algoritmo en tanto que secuencia de instrucciones que constituyen el motor de funcionamiento de cualquier programa informático y por ende quedaría protegido por los derechos de autor, y para otro sector doctrinal no estaría protegido por esta vía en tanto en cuanto la *Directiva 2009/24/CE sobre la protección jurídica de los programas de ordenador*, en su Considerando 11, aclara que solo se protege la expresión del programa de ordenador pero no *las ideas y principios implícitos en los elementos del programa, incluidas las de sus interfaces*, disposición que después se recoge en su art. 1.2 y que también establece el art. 96.4 LPI. En este último sentido, ALAMEDA CASTILLO, M.T., *op.cit.* p. 34 o CASTILLO PARRILLA, J.A., «Sentencia del Tribunal Ordinario de Bolonia de 31 de diciembre de 2020 (Caso Deliveroo) ¿Discriminación algorítmica o discriminación a través de un algoritmo?», *Derecho Digital e Innovación. Digital Law and Innovation Review*, n.7 (octubre-diciembre), 2020, para quien la protección jurídica de los algoritmos se reduce a su protección como secretos comerciales, ya que no pueden ser objeto de derechos de autor ni son patentables. Defendiendo la primera postura, RIVAS VALLEJO, P. «Análisis…», *cit.*, p. 442, citando también la sentencia del TJUE de 22 de diciembre de 2010 (asunto *Bezpecnostni softwarová asociase)*, favorable a la protección del algoritmo en base a la Directiva 2009/24/CE.

[37] Se trata de un *software,* de titularidad pública pero gestionado por empresas privadas, que determina qué consumidores en situación de vulnerabilidad tienen derecho a un descuento en la factura de la luz. La concesión de este beneficio no es una decisión discrecional de la Administración, sino que las condiciones para obtenerlo y la forma de solicitarlo están previstas en la ley. Se regula en el Real Decreto 897/2017 (desarrollado por la Orden ETU/943/2017) que aplica una tarifa eléctrica reducida a consumidores vulnerables. Sobre este caso, puede verse, entre otros, BAZ LOMBA, C. (2021). Los algoritmos y la toma de decisiones administrativas. Especial referencia a la transparencia. *Revista CEFLegal,* 243, 119-160; también COTINO HUESO, L., «Caso Bosco, a la tercera tampoco va la vencida. Mal camino en el acceso los algoritmos públicos», Diario LA LEY, Nº 84, Sección Ciberderecho, 17 de Mayo de 2024.

[38] En principio, la aplicación debía ser neutral porque solo tenía que realizar la mera aplicación de los requisitos de acceso a la prestación sin cambiar nada del panorama legal ni práctico. Sin embargo, tras su implantación, la mitad de los anteriores beneficiarios dejaron de ser idóneos para el reconocimiento del bono social. De aquí se infiere que, aunque en principio las reglas del juego sean idénticas, la forma de trabajar de los algoritmos que eleva exponencialmente el índice de gra-

Tanto el Ministerio para la Transición Ecológica, como el CTBG y los diferentes órganos judiciales que conocieron el asunto, denegaron a CIVIO el acceso al código fuente basándose en que afectaría no sólo a los derechos de propiedad intelectual sobre el *software*, sino también a la seguridad pública y a la defensa nacional por la conexión del sistema con bases de datos de carácter sensible cuya integridad podía verse comprometida[39]. La Sala de lo Contencioso-Administrativo de la Audiencia Nacional (sección séptima), en su reciente sentencia de 30 de abril de 2024 (que rechaza por tercera vez la petición de acceso al código fuente), reitera que el mismo está protegido por la Ley de Propiedad Intelectual y que carece de todo fundamento excluir a la Administración del derecho de la protección intelectual. Además, considera que la difusión del código fuente de la aplicación *Bosco* pondría en grave riesgo derechos de terceros y atentaría a bienes jurídicos protegidos por los límites al derecho de acceso a la información pública del artículo 14.1 letras d), g), i) y k). Concluye que, sin lugar a dudas, la ocultación del código fuente de la aplicación es garantía de protección ante las vulnerabilidades[40].

Aunque éste no es el único caso sobre decisiones algorítmicas que ha sido objeto de resolución judicial por nuestros tribunales y órganos administrati-

---

nularidad del análisis de datos disponibles, cambia los resultados. Así lo explica RIVAS VALLEJO, P., «Sesgos de automatización y discriminación algorítmica», en *Discriminación algorítmica en el ámbito laboral*, Pilar Rivas Vallejo (dir.), 2022, pp. 52-53.

[39] Nos parecen interesantes las consideraciones que realiza el CTBG sobre la protección jurídica del código fuente: «El código fuente es el archivo o conjunto de archivos que tienen un conjunto de instrucciones muy precisas, basadas en un lenguaje de programación, que se utiliza para poder compilar los diferentes programas informáticos que lo utilizan y se puedan ejecutar sin mayores problemas […]. El *software* ha sido extraordinariamente difícil de clasificar como materia específica de propiedad intelectual debido a que su doble naturaleza plantea problemas particulares para quienes tratan de establecer analogías con las categorías jurídicas existentes. Esta es la razón por la que ha habido intentos de clasificarlo como objeto de derechos de autor, de patentes o de secretos comerciales, e incluso como un derecho *sui géneris* de *software* […] Éste es el enfoque vigente respecto de la protección del software en diversos tratados internacionales. Así, por ejemplo, el artículo 4 del Tratado de la OMPI sobre Derecho de Autor (WCT), el artículo 10 del Acuerdo sobre los ADPIC de la Organización Mundial del Comercio y el artículo 1 de la Directiva (91/250/CEE) del Consejo Europeo sobre la protección jurídica de programas de ordenador equiparan el software con las obras literarias, protegidas por el derecho de autor. A efectos de la Directiva, el término «programa de ordenador» incluye programas en cualquier forma, incluso los que están incorporados en el «hardware»; este término designa también el trabajo preparatorio de concepción que conduce al desarrollo de un programa de ordenador, siempre que la naturaleza del trabajo preparatorio sea tal que más tarde pueda originar un programa de ordenador».

[40] En opinión de COTINO HUESO, L., «Caso Bosco…», *cit.*, quien critica la calidad técnica de la resolución, «lo que se alega -y triunfa- *por la puerta de atrás* es la seguridad informática, que no había sido alegada convenientemente por la Administración cuando tocaba». Los tribunales se basan en diversos informes técnicos para mantener que revelar el código fuente podría comprometer la seguridad de la información y aumentar la vulnerabilidad ante ataques informáticos, lo que podría afectar adversamente la integridad de datos personales y la seguridad pública.

vos, ha sido uno de los más relevantes y criticados por entender la doctrina que la solicitud de transparencia de la Fundación CIVIO estaba más que justificada, ya que no puede tolerarse que se adopten decisiones desestimatorias de beneficios aprobados por el Gobierno en favor de determinados colectivos sin conocer cómo se llega a ellas, lo que genera indefensión y un enorme riesgo de arbitrariedad de los entes públicos[41]. Tanto el CTBG estatal como la GAIP, se han pronunciado en diversas ocasiones sobre el acceso al código fuente llegando la mayoría de las veces (aunque no todas) a soluciones favorables al interesado que después han sido anuladas por los órganos jurisdiccionales[42]. Así ocurrió, por ejemplo, en la resolución RT/253/2021 que fue recurrida por la Comunidad de Madrid y el recurso contencioso-administrativo fue estimado por la Sentencia nº 158/2022 del Juzgado Central Contencioso-Administrativo nº 11[43].

---

[41] Criticando la postura adoptada por los distintos órganos y tribunales en este caso, puede verse, entre otros, DE LA NUEZ SÁNCHEZ CASCADO, E., «Algoritmos y transparencia», disponible en https://www.hayderecho.com/2020/02/18/algoritmos-y-transparencia-2/, para quien, lo mínimo que se debería exigir es conocer cómo funciona el sistema algorítmico y cómo adopta decisiones que son de enorme relevancia para personas en situación o riesgo de pobreza energética. Para PONCE SOLÉ, J., «A propósito de la sentencia de la Audiencia Nacional de 30 de abril de 2024, núm. de rec. 51/2022, sobre el caso BOSCO», recuperado de Por qué se equivocan las sentencias sobre el algoritmo del bono social eléctrico - Almacén de Derecho (almacendederecho.org), los riesgos aducidos en la sentencia de la Audiencia Nacional conllevarían que pudiera denegarse en cualquier caso el acceso al código fuente, postura que se separa de la que están adoptando otros tribunales de nuestro entorno y órganos administrativos como el CTBG. Criticando también esta postura, COTINO HUESO, L., «Caso Bosco…» *cit.*, y HUERGO LORA, A. (2024) «Por qué aciertan las sentencias sobre el 'algoritmo' del bono social eléctrico», accesible en https://almacendederecho.org/por-que-aciertan-las-sentencias-sobre-el-algoritmo-del-bono-social-electrico (último acceso 31 de julio de 2024).

[42] Pueden verse la Resolución de 21 de septiembre de 2016 de la Comisión de Garantía del Derecho de Acceso a la Información Pública (GAIP), que estima las Reclamaciones 123/2016 i 124/2016 (acumuladas) sobre el algoritmo matemático que determina la selección de los miembros de los tribunales correctores de las pruebas de acceso a la Universidad (PAU) y DNI de los candidatos presentados, y su Resolución 200/2017, de 21 de junio relacionada con las anteriores. Por lo que se refiere al CTBG, las resoluciones R/0058/2021 de 20 de mayo, sobre el acceso al algoritmo para calcular pensiones de la TGSS; RT/253/2021 de 19 de noviembre y RT/ 748/2021 de 10 de enero 2022 sobre acceso al código fuente de aplicación utilizada para el sorteo de tribunales asociados a procesos selectivos para educación en Madrid, y la Resolución del Expediente 551-2023 sobre el acceso a información del Sistema de Seguimiento Integral de los casos de Violencia de Género (Sistema VioGén).

[43] El juzgado confirma la aplicación del límite del artículo 14.1.j) de la LTAIBG alegado por la Comunidad para denegar el acceso al código fuente y considera que la petición del solicitante puede considerarse abusiva, vinculando el abuso a su carácter excesivo e injustificado, que excedería del contenido razonable del derecho de acceso a una información pública cabalmente considerado, teniendo en cuenta que el administrado tuvo conocimiento, en todo caso, del funcionamiento del sistema para la designación de los vocales de los tribunales de procesos selectivos cuyo resultado y funcionamiento es público. Por tanto, concluye que en el caso concreto se había dado transparencia al procedimiento público y que los interesados tuvieron la posibilidad de acudir al sorteo público celebrado, teniendo acceso a los resultados de los distintos sorteos al ser públicos.

En relación con el acceso a dicho código, el art. 74.13 RIA aprobado en fechas muy recientes, posibilita que las *autoridades de vigilancia del mercado* (no el público en general) puedan tener acceso al código fuente de los sistemas de alto riesgo, previa solicitud motivada, si se cumplen dos condiciones: a) que dicho acceso sea necesario para llevar a cabo la evaluación de la conformidad del sistema con los requisitos establecidos en el capítulo III, sección 2; y b) que se hubieran agotado todos los procedimientos de prueba o auditoría y todas las comprobaciones basadas en los datos y la documentación facilitados por el proveedor, o estos fueran insuficientes. Ahora bien, como establece el apartado 14, toda la información o documentación a la que tengan acceso tales autoridades de vigilancia deberá ser tratada de conformidad con las obligaciones de confidencialidad establecidas en el artículo 78. Ello con el fin de proteger los derechos de propiedad intelectual e industrial y la información empresarial confidencial o los secretos comerciales de una persona física o jurídica, incluido el código fuente.

Por último, otro escollo a la transparencia que a su vez dificultaría la averiguación del origen de la discriminación algorítmica, sería la propiedad de los datos utilizados para entrenar al algoritmo[44]. Esta cuestión no es baladí por cuanto los expertos en *machine learning* aseveran que lo realmente útil para detectar el sesgo del sistema podría ser el acceso a los conjuntos de datos que han servido de soporte al aprendizaje del algoritmo[45]. Esto ocurriría en los casos de algoritmos predictivos cuyo aprendizaje se basa precisamente en los datos aportados, a diferencia de lo que ocurriría en los algoritmos no predictivos en los que los datos de alimentación son secundarios; en estos últimos, sí que podría ser útil acceder al código fuente o a la secuencia de programación para averiguar el origen de la decisión, detectando así fallos o errores en alguno de sus pasos[46].

Hasta hace poco tiempo ninguna norma del acervo comunitario obligaba expresamente a ofrecer acceso a los datos de entrenamiento para investigar posibles errores del sistemas o sesgos discriminatorios[47]. Para López

---

[44] RIVAS VALLEJO, P. «Análisis…», *cit.,* pp. 448 y ss. Sobre el derecho de acceso a los datos de entrenamiento de sistema, puede verse ampliamente, LÓPEZ-TARRUELLA MARTÍNEZ, A., *Propiedad intelectual e innovación basada en los datos*, Navarra, 2021, pp. 192 y ss.

[45] RIVAS VALLEJO, P., «Herramientas desde el derecho antidiscriminatorio y la protección frente a la igualdad y a la no discriminación», p. 548.

[46] RIVAS VALLEJO, P. «Análisis …», *cit.,* p. 439.

[47] Para LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.*, p. 199, esta laguna podía haberse cubierto mediante una clarificación jurisprudencial o legislativa del derecho de explicabilidad y de su alcance. También considera que podría invocarse el art. 1.2 c) de la Directiva 2016/943 para impedir que el titular de los datos hiciera valer la protección por secreto empresarial para negarse a conceder acceso, al menos en aquellos supuestos en los que el solicitante fuera una autoridad pública. En el acervo comunitario, la fallida *Propuesta de Reglamento sobre un régimen de responsabilidad civil en materia*

Tarruella, cuya opinión compartimos, resultaba paradójico que no existiera en el Derecho de la UE una norma de propiedad intelectual que facilitara la investigación de sesgos algorítmicos en general, o de aquellos que pudieran afectar a derechos fundamentales en particular, sobre todo teniendo en cuenta que para las instituciones europeas la transparencia y la explicabilidad eran componentes esenciales de la IA centrada en el ser humano. Para dicho autor, en estos supuestos estaría justificado el acceso a los datos del sistema, e incluso podría considerarse una cuestión de orden público, en tanto que averiguar si la IA discrimina no solo interesa a los particulares afectados sino también a los Gobiernos[48]. Ahora bien, dicho acceso habría que concederlo bajo ciertas salvaguardas que garantizasen, por un lado, los legítimos intereses del titular de los datos, y por otro lado, la protección de la información confidencial[49].

Pues bien, en fechas recientes se han aprobado diversas normas que garantizan un derecho de acceso a los datos e información algorítmica. Baste citar por el momento, a modo de ejemplo, que la nueva *Ley de Datos* que entró en vigor el 11 de enero de 2024 y será aplicable en septiembre de 2025[50], establece en su art. 4 el derecho del usuario de un producto o servicio digital contratado a acceder a los datos que el mismo genere. En opinión de Navas

---

de inteligencia artificial, de 20 de octubre de 2020, sí que contemplaba un derecho de acceso a los datos generados por el sistema de alto riesgo (fuesen datos personales o de otro tipo, especialmente técnicos) tanto a la víctima como al operador responsable. Sobre esto el art. 10.2 disponía lo siguiente: «Un operador considerado responsable podrá utilizar los datos generados por el sistema de IA para demostrar la negligencia concurrente de la persona afectada, de conformidad con el Reglamento (UE) 2016/679 y otras leyes en materia de protección de datos relevantes. La persona afectada también podrá usar esos datos con fines probatorios o aclaratorios en la demanda por responsabilidad civil».

[48] En este sentido, LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.*, p. 195. Precisamente ese fundamento es el esgrimido en el RIA para justificar el tratamiento de categorías especiales de datos a través de sistemas de IA. Así, evitar la discriminación que podría provocar el sesgo de los sistemas de IA se considera una cuestión de orden público esencial que permitiría que los proveedores trataran ese tipo de datos. De ahí que su art. 10.5 disponga que «*En la medida en que sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo de conformidad con lo dispuesto en el apartado 2, letras f) y g), del presente artículo, los proveedores de dichos sistemas podrán tratar excepcionalmente las categorías especiales de datos personales siempre que ofrezcan las garantías adecuadas en relación con los derechos y las libertades fundamentales de las personas físicas*». Para que se produzca dicho tratamiento deberán cumplirse, además, las condiciones que se establecen en el precepto junto con las disposiciones establecidas en los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y la Directiva (UE) 2016/680.

[49] LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.*, pp. 195-196. Considera el autor que regulando el acceso a los datos bajo ciertas condiciones se evitaría desincentivar a las organizaciones a invertir en el desarrollo de herramientas de IA ante la posibilidad de verse obligados a compartir los conjuntos de datos en los que se basa su ventaja competitiva. Vid., p. 201.

[50] Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo, de 13 de diciembre de 2023, sobre normas armonizadas para un acceso justo a los datos y su utilización, y por el que se modifican el Reglamento (UE) 2017/2394 y la Directiva (UE) 2020/1828 (Reglamento de Datos).

Navarro, dicha norma ampara el acceso a los datos a los efectos de una posible futura litigación por daños[51].

No obstante, son las últimas *Propuestas de Directivas de la Comisión Europea*, de 28 de septiembre de 2022, sobre *responsabilidad extracontractual en la IA* y sobre *modernización del régimen de responsabilidad por productos defectuosos* (en adelante, PDRC y PDPD respectivamente)[52], las que han introducido importantes medidas a fin de remover los obstáculos para acceder a la información algorítmica mediante el reconocimiento de un *derecho a la exhibición de pruebas* a favor de las víctimas. A pesar de su novedad y relevancia, no podemos detenernos en el análisis de dichas normas en este trabajo pues el mismo se centra en otras medidas legislativas. Simplemente apuntaremos que, a nuestro modo de ver, con dicho expediente la Comisión trató de suplir la carencia inicial del RIA respecto a la transparencia externa hacia los afectados por las decisiones de los sistemas de IA, estableciendo un derecho de acceso a la información y a los datos en poder de demandados o potenciales demandados (*vid.,* arts. 3 y 8 respectivamente). El objetivo de esta medida era eliminar los obstáculos con los que podían encontrarse las víctimas para acceder a información del sistema que le permitiera probar los presupuestos para reclamar responsabilidad civil por los daños sufridos e incentivar a proveedores y responsables del despliegue de los sistemas a facilitar información sobre estos y a cumplir las medidas del RIA dirigidas a garantizar la transparencia.

3. TRANSPARENCIA ALGORÍTMICA Y EXPLICABILIDAD: ¿QUÉ IMPLICAN ESTAS EXIGENCIAS?

No es suficiente con requerir que los algoritmos o los sistemas de IA sean transparentes o explicables, sino que hay que determinar qué implican estas exigencias a lo largo de su ciclo de vida y cómo se controlará su cumplimiento.

Aunque en ocasiones ambos términos se emplean como sinónimos, hay que advertir que se trata de dos características distintas de los sistemas de IA, si bien íntimamente conectadas.

Por lo que se refiere concretamente a la transparencia, se ha precisado que es importante distinguir también entre la *transparencia en el ámbito de la IA* y la *transparencia algorítmica*, ya que, mientras la primera se refiere a la «capacidad de entender cómo funcionan los sistemas de IA en general, cómo se

---

[51] NAVAS NAVARRO, S., *Daños ocasionados por sistemas de inteligencia artificial. Especial atención a su futura regulación,* Granada, 2022, p. 71.

[52] COM (2022) 496 final (2022/0303(COD)) y COM(2022) 495 final (2022/0302(COD)).

entrenan, cómo toman decisiones y cómo afectan a los usuarios y a la sociedad en su conjunto», la segunda «se centra específicamente en hacer visibles los factores que influyen en las decisiones tomadas por los algoritmos, de modo que puedan ser comprendidos por las personas que utilizan, regulan y se ven afectadas por estos sistemas, siendo su objetivo fundamental garantizar la exactitud y la equidad en las decisiones que afectan a las personas»[53].

Según establece la *Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica,* el principio de transparencia algorítmica «consiste en que siempre ha de ser posible justificar cualquier decisión que se haya adoptado con ayuda de la IA y que pueda tener un impacto significativo sobre la vida de una o varias personas». Es decir, «siempre debe ser posible traducir los cálculos del sistema de inteligencia artificial a una forma comprensible para los humanos», matizándose en otros instrumentos de *soft law* que dicho principio «no consiste en revelar códigos, sino en hacer inteligibles los parámetros y criterios de las decisiones que se toman»[54]. Su finalidad principal es que los usuarios de los sistemas puedan entender su funcionamiento y resultados.

Según el Estudio de Impacto algorítmico del Ministerio de Interior de Países Bajos, la transparencia puede ser tanto *interna* como *externa*. La primera va dirigida principalmente a los sujetos de la cadena de valor de los sistemas (proveedores, responsables del despliegue, importadores, distribuidores, etc.) en beneficio de auditores, supervisores, autoridades o jueces, y es esencial para conocer y comprender su funcionamiento y detectar posibles errores[55]; la segunda, va dirigida al público en general o al menos a todos los

---

[53] FERNÁNDEZ ROZAS, J.C., «La Ley de Inteligencia Artificial de la Unión Europea: un modelo para innovaciones radicales, responsables y transparentes basadas en el riesgo», LA LEY Unión Europea, Nº 124, Sección Estudios, Abril 2024, LA LEY 15196/2024, p. 24/59. Recuperado de https://www.congreso.es/docu/docum/ddocum/notasdocumentales/nd1/inteligencia_artificial.pdf.

[54] Así, el Dictamen del CESE sobre *Inteligencia artificial: anticipar su impacto en el trabajo para garantizar una transición justa;* la Recomendación del Consejo OCDE sobre IA de 2019 también recoge entre sus principios la exigencia de que los sistemas de IA sean transparentes y explicables lo que implica proporcionar información significativa, adecuada al contexto y coherente; hacer comprensible los sistemas de IA, para que las partes interesadas sean conscientes de sus interacciones con estas herramientas, especialmente en el lugar de trabajo; permitir que los afectados por un sistema de inteligencia artificial entiendan su resultado y puedan cuestionarlo. Vid., p.5/7. Recuperado de: file:///C:/Users/ASUS/Downloads/db5053b5-93e0-4cf5-a7cf-edce5ee6e893.pdf. Según el Estándar de Transparencia Algorítmica desarrollado por la Oficina Central Digital y de Datos de Reino Unido, «la transparencia algorítmica significa dar a conocer cómo las herramientas algorítmicas apoyan las decisiones, incluyendo proporcionar información sobre dichas herramientas y las decisiones asistidas por algoritmos en un formato completo, abierto, comprensible, de fácil acceso y gratuito».

[55] A esta transparencia interna o comunicación de la información entre los distintos sujetos de la cadena de valor de los sistemas se le otorga una especial relevancia a lo largo del RIA, como

interesados o afectados por la IA y también se conoce como «explicabilidad pública»[56].

La transparencia guarda una estrecha relación con el *principio de explicabilidad*. Así se reconoce, por ejemplo, en la *Recomendación de la UNESCO sobre la ética de la IA de noviembre de 2021*, donde se afirma que ambos conceptos están íntimamente relacionados entre sí, y con la responsabilidad, la rendición de cuentas y la confianza en los sistemas. Además, suelen ser condiciones previas fundamentales para garantizar el respeto, la protección y la promoción de los derechos humanos, las libertades fundamentales y los principios éticos. No obstante, como se explica en dicho documento, mientras la *transparencia* pretende proporcionar información para que las personas puedan comprender cómo se implementa cada etapa de un sistema de IA en función de su contexto, conocer los factores que influyen en una predicción o decisión específicas y si existen o no medidas de seguridad o de equidad (e incluso excepcionalmente, en casos de amenazas graves con repercusiones adversas para los derechos humanos, podría exigir que se compartiera el código fuente o los conjuntos de datos utilizados), la *explicabilidad* tiene como objetivo hacer inteligible la entrada, la salida y el funcionamiento de cada componente algorítmico y la forma en que contribuye a los resultados del sistema[57].

Esta conexión entre la transparencia y la explicabilidad también la pone de relieve el Grupo de Expertos en sus *Directrices éticas para una IA fiable*, donde, además de consagrarse la explicabilidad como uno de los cuatro *principios éticos* que deben respetarse en el desarrollo, despliegue y utilización de los sistemas de IA para garantizar los derechos fundamentales, se incluye como componente fundamental de la transparencia junto con la trazabilidad y la comunicación[58]. No obstante, como aclara el Considerando 49, pese a que

---

después se explicará. También es a la que se refieren los organismos de normalización técnica NIST (EEUU), ISO, CEN CENELEC (UE).

[56] Así lo explica COTINO HUESO, basándose en el Estudio citado. Vid., COTINO HUESO, L., «Transparencia de la inteligencia artificial pública: marco legal, desafíos y propuestas», *Actualidad Administrativa, Nº I, Sección Actualidad,* Diciembre 2023. Recuperado de La Ley Digital, LA LEY 12767/2023. Para este autor, la transparencia y la explicabilidad permiten comprender resolver problemas técnicos del funcionamiento del sistema, especialmente para comprender la cadena de causalidades.

[57] Vid., ap.37 a 41.

[58] *Directrices éticas,* p. 22. COTINO HUESO, L., «Transparencia de la inteligencia artificial pública…», cit., p. 3/23, califica el concepto de transparencia que proporciona el Grupo de expertos en IA de la Comisión Europea en sus *Directrices éticas* como un «concepto inclusivo». Nótese que dicho concepto de transparencia mantenido por el Grupo de Expertos en IA en sus Directrices éticas, ha sido el acogido en la versión final del RIA. Así, en el Considerando 27 se establece que, de acuerdo con las citadas Directrices, por transparencia se entiende que «los sistemas de IA se desarrollan y utilizan de un modo que permita una *trazabilidad y explicabilidad* adecuadas, y que, al mismo tiempo, *haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente*

numerosas obligaciones legales reflejan principios éticos, el cumplimiento de estos últimos trasciende el mero cumplimiento de las leyes existentes.

Como *principio rector*, se entiende que la explicabilidad «es crucial para conseguir que los usuarios confíen en los sistemas de IA y para mantener dicha confianza» e implica que «los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de IA y que las decisiones deben poder explicarse -en la medida de lo posible- a las partes que se vean afectadas por ellas de manera directa o indirecta», ya que, si no fuera así, se les estaría privando de la posibilidad de presentar alegaciones e impugnarlas adecuadamente[59].

Como *elemento integrante de la transparencia*, se define como la «capacidad de explicar tanto los procesos técnicos de un sistema de IA, como las decisiones humanas asociadas» y se establece que dicha exigencia requiere que las decisiones del sistema «sean comprensibles para los seres humanos y estos tengan la posibilidad de rastrearlas»[60].

El *grado* de explicabilidad exigible va a depender en gran medida del contexto en el que se enmarquen estos sistemas y de la gravedad de las consecuencias derivadas de un resultado erróneo o impreciso (por ejemplo, un sistema de IA que genere unas recomendaciones de compra poco acertadas no será preocupante desde el punto de vista ético, a diferencia de uno dirigido a evaluar la concesión de la libertad condicional al condenado) y debe ser *suficiente* para poder impugnar los resultados y solicitar un resarcimiento por los daños que se generen[61]. Además, en la medida de lo posible, la explicabilidad debe estar adaptada a las personas afectadas[62].

---

a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos». Este Considerando 27 contiene lo dispuesto en la enmienda 213 del Parlamento europeo que proponía un nuevo art. 4 bis donde se recogieran los «Principios generales aplicables a todos los sistemas de IA», en cuyo apartado 1. d) figuraba la transparencia.

59 Vid., ap. 53 *Directrices éticas.*

60 P. 22.

61 Así lo afirma el Parlamento Europeo en la Propuesta de Reglamento que acompañaba a su Resolución sobre aspectos éticos de la IA, de 20 de octubre de 2020. Vid., Considerando 19, donde finaliza diciendo que «La auditabilidad, la trazabilidad y la transparencia deben tratar cualquier posible ininteligibilidad de estas tecnologías». La UNESCO en su *Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO* (2021), señala que dicho grado de transparencia deberá ajustarse al contexto y efecto del sistema de IA.

62 Vid., Comunicación de la Comisión Europea *«Generar confianza en la inteligencia artificial centrada en el ser humano»* (COM/2019/168 final), p. 6 o *Directrices éticas* del Grupo de Expertos, p. 22, donde se apunta que la explicación «debería ser oportuna y adaptarse al nivel de especialización de la parte interesada (que puede ser una persona no experta en la materia, un regulador o un investigador)».

Para Cotino Hueso, las variables que podrían condicionar el grado o intensidad de explicabilidad y transparencia exigible serían las siguientes: en primer lugar, el impacto que tiene el algoritmo en la decisión, el resultado y el ciudadano; en segundo lugar, el grado de autonomía del sistema en la toma de decisiones (es decir, hasta qué punto se garantiza la participación humana); y por último, el tipo y la complejidad del algoritmo. Así, cuanto mayor sea el impacto de las decisiones algorítmicas en los ciudadanos y mayor sea la autonomía, opacidad y complejidad del sistema de IA, mayores niveles de transparencia y explicabilidad se tendrán que exigir[63].

Por lo que se refiere a los algoritmos de *black box*, en los que no es posible obtener una explicación acerca de cómo se ha generado un resultado determinado o qué combinación de *inputs* han contribuido al mismo, la explicabilidad debe traducirse en medidas como la trazabilidad, la auditabilidad o la comunicación transparente de las capacidades del sistema para que las personas afectadas puedan comprender cuál es la base de sus acciones o decisiones[64]. Es decir, en tales casos es necesario adoptar otro tipo de estrategias de control con el fin de desarrollar sistemas algorítmicos explicables desde un punto de vista práctico. Más allá de la apertura de la caja negra (que como hemos visto no siempre será la panacea para la víctima), la doctrina ha venido manteniendo que dichas estrategias debían ir dirigidas al escrutinio de los datos de entrada y salida y al establecimiento de una serie de garantías en el proceso de diseño, entrenamiento e implementación de los sistemas y de procedimientos de monitorización y supervisión de los resultados. Garantías y procedimientos que, además, debían someterse a auditorías[65]. Como iremos viendo a lo largo de estas páginas, muchas de estas estrategias han sido recogidas en el nuevo Reglamento europeo sobre IA.

---

[63] Así lo entiende COTINO HUESO, L., «Transparencia de la inteligencia artificial pública…», cit. También en «Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida» *Revista Española de la Transparencia,* Núm. 16, 2023, pp.17-63, accesible en file:///C:/Users/ASUS/Downloads/Dialnet-QueConcretaTransparenciaEInformacionDeAlgoritmosEI-8913030.pdf

[64] La Comisión Europea ha señalado en diversos documentos que en relación a aquellas aplicaciones de IA que entrañan un riesgo elevado, el marco regulador podrá exigir la conservación de registros exactos sobre el conjunto de datos utilizados para entrenar el algoritmo, una descripción de sus principales características y del modo en que se escogió, así como de documentación sobre las metodologías de programación y entrenamiento y los procesos y las técnicas utilizados para construir, probar y validar los sistemas; además de obligar a facilitar información clara de sus capacidades y limitaciones, objetivo y condiciones en las que se espera que funcione según lo previsto. Incluso considera el diseño de *sistemas mixtos* que combinen redes neuronales profundas y razonamiento simbólico (esto es, reglas creadas mediante intervención humana) como una solución técnica que ayude a mejorar la capacidad de explicar los resultados de la IA. Vid., *Directrices éticas* p. 16 y también el *Libro Blanco sobre la inteligencia artificial* (COM (2020) 65 final), pp. 23-25.

[65] Por todos, SOLAR CAYÓN, J.I., *op.cit.,* p. 152.

Además de la explicabilidad, la transparencia incluye también la *trazabilidad* y la *comunicación*[66]. Esto significa que no sólo se debe poder reconstruir cómo y por qué un sistema se comporta de determinada manera, sino que quienes interactúen con ellos deben saber que no se trata de un humano sino de un sistema inteligente y qué personas responderán de sus acciones o decisiones.

A mayor abundamiento, la transparencia también debe aplicarse a los elementos del sistema de IA, es decir, los datos, el *software* y los modelos de negocio[67]. Para conseguirla, deben documentarse los conjuntos de datos y procesos que dan lugar a la decisión del sistema, incluidos los relativos a la recopilación y etiquetado de datos, estar disponibles las explicaciones sobre el grado en que un sistema de IA condiciona e influye en el proceso de toma de decisiones de la organización, las opciones de diseño de dicho sistema y la justificación de su despliegue[68]. En relación a este último extremo, algunos estudios advierten que puede haber otros problemas de transparencia que no estén relacionados con el algoritmo en sí mismo o con los datos que utiliza, sino con la falta de claridad en cuanto a las razones por las que se usa un determinado sistema de IA y los objetivos que persigue. En Francia, por ejemplo, el algoritmo *Parcoursup*, que se utilizaba para asignar a los estudiantes a las universidades, fue muy criticado porque sus objetivos no estaban bien definidos[69].

Respecto a los datos, cuando estos sistemas empleen datos *personales* será de obligada aplicación el RGPD, en cuyo contexto la transparencia se configura como una obligación para los responsables del tratamiento a la hora de comunicar determinada información al interesado, lo que debe hacerse de forma «concisa, transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo» según dispone su art. 12.1. De ahí que se infiera que la principal preocupación del RGPD es la transparencia «hacia la persona interesada», más allá de las posibles soluciones técnicas para ayudar a esclarecer lo que hasta el momento es inescrutable[70].

---

[66] Sobre ambas características, vid., *Directrices éticas,* p. 22.

[67] *Ibidem.* También la Agencia Europea de los Derechos Fundamentales mantiene que, tanto la transparencia como las evaluaciones de impacto, deben verificarse tanto sobre la calidad de los datos, como sobre la forma en que funcionan los algoritmos (caja negra) y las facultades predictivas de los mismos. Vid., EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (FRA), «BigData: Discrimination in data-supported decision making» 2018.

[68] Vid., Comunicación de la Comisión Europea *«Generar confianza en la inteligencia artificial centrada en el ser humano»* (COM/2019/168 final), p. 6.

[69] GERARDS, J./ XENIDIS, *op.cit*, p. 87.

[70] LAZCOZ MORATINOS, G., *op.cit.,* p. 296.

Sin embargo, como apunta la AEPD, en los casos de *tratamientos de datos basados en IA* (esto es, cuando los sistemas de IA se incluyen en, o son medios de, un tratamiento de datos personales) «la transparencia debe permitir a los interesados ser conscientes del impacto que el empleo de dichas soluciones puede llevar asociado y de ahí que la transparencia esté dirigida tanto a los interesados como a los operadores del tratamiento, y que esté ligada, en particular, con una *información veraz sobre la eficiencia, las capacidades y las limitaciones reales de los sistemas de IA, que evite la creación de falsas expectativas, en los usuarios y los interesados, que puedan ocasionar una mala interpretación de las inferencias que se realizan en el marco del tratamiento. También, la transparencia está ligada a información sobre el contexto y situación del tratamiento, como la existencia de terceras partes, la localización física/virtual de la solución AI,* etc». En tales casos, los responsables del tratamiento deben obtener suficiente información sobre los sistemas para cumplir con las obligaciones que les impone el RGPD y que incluyen la transparencia para permitir el ejercicio de los derechos, el cumplimiento del principio de responsabilidad activa, o cumplir los requisitos de las Autoridades de Supervisión en relación con sus poderes de investigación y de los organismos de certificación y supervisión de códigos de conducta. Además, como sigue diciendo la AEPD, «[l]a transparencia no se reduce a un instante puntual, sino que debe ser entendida como un principio en torno al que orbita de forma dinámica el tratamiento realizado y que afecta a todos y cada uno de los elementos y participantes que intervienen en la solución»[71].

Por tanto, es importante percatarse ya de entrada de que la transparencia tiene diferentes implicaciones en el RIA y en el RGPD en la medida en que afecta a distintos actores, se refiere a información diferente y va dirigida a diferentes destinatarios[72]. Como después se detallará, en el RIA la transparencia es exigible desde el diseño del sistema y a lo largo de todo su ciclo de vida independientemente de si el mismo procesa o no datos personales. Va referida a los sistemas en sí mismos y las obligaciones que de ella derivan van dirigidas a diseñadores, desarrolladores, proveedores y responsables del despliegue. Por su parte, la transparencia recogida en el RGPD se aplica a los tratamientos de datos personales definidos en los arts. 2 y 4.2 y las obligaciones que implica recaen sobre los responsables de dicho tratamiento. No obstante, como aclara la AEPD, «los *diseñadores* y *desarrolladores* pueden ser responsables o encargados del tratamiento si utilizan datos personales en el diseño o desa-

---

[71] Informe de la AEPD sobre la *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*, 2020, p. 33

[72] Como explica la AEPD en https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia, el RIA (cuyo ámbito material son los sistemas de IA) establece un concepto de transparencia que difiere del mismo término establecido en el RGPD (cuyo ámbito material son los tratamientos de datos personales).

rrollo del sistema de IA. Los *proveedores* pueden ser responsables o encargados del tratamiento si los sistemas de IA almacenan o tratan datos de interesados identificados o identificables. [Y] los *usuarios/entidades que despliegan* un sistema de IA podrían ser responsables o encargados del tratamiento si incluyen dicho sistema como parte de sus tratamientos»[73]. En cualquier caso, sólo aquellos que actúen como responsables del tratamiento quedarán sujetos a las obligaciones de transparencia del RGPD.

Por último, cabe señalar también que la transparencia a la que se refiere el RIA se dirige principalmente a los responsables del despliegue y está relacionada con la explicabilidad, documentación, mantenimiento de registros e información sobre cómo utilizar dicho sistema de IA[74], mientras que los destinatarios de las obligaciones de transparencia derivadas del RGPD son los interesados -personas físicas- y su contenido mínimo se recoge en los arts. 13 y 14 RGPD. No obstante, como también se explicará con detenimiento después, aunque el RIA abogue principalmente por la transparencia interna, también establece obligaciones a favor de la transparencia externa y ha acabado incorporando, como novedad relevante respecto a su versión inicial, el *derecho a una explicación* de las personas afectadas por las decisiones algorítmicas.

4. MEDIDAS PARA GARANTIZAR LA TRANSPARENCIA Y LA EXPLICABILIDAD EN LA TOMA DE DECISIONES ALGORÍTMICAS

4.1 Estado de la cuestión

Hasta hace poco tiempo el único instrumento para hacer frente a la opacidad de los sistemas automatizados de decisiones, basados o no en IA, era el RGPD. Sin embargo, en los últimos años se han presentado otras propuestas

---

[73] https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia. Sobre esto, el Considerando 10 RIA señala que el Reglamento no afectará a las obligaciones de los proveedores y responsables del despliegue en su papel de responsables o encargados del tratamiento si el diseño, desarrollo o uso del sistema IA implica tratamiento de datos personales. Las normas armonizadas del RIA deberán facilitar la aplicación efectiva y permitir el ejercicio de los derechos y otras vías de recurso de los interesados garantizados por el Derecho de la Unión en materia de protección de datos personales, así como de otros derechos fundamentales.

[74] Vid., por ejemplo, el art. 13 y el Considerando 72, donde se da amplia cuenta de que la transparencia exigible a los sistemas de IA de alto riesgo va dirigida, principalmente, a que los responsables del despliegue puedan «comprender la manera en que el sistema de IA funciona, evaluar su funcionalidad y comprender sus fortalezas y limitaciones», así como «utilizar el sistema y tomar decisiones con conocimiento de causa».

regulatorias por las instituciones europeas y se han ido aprobado paulatinamente normas a nivel interno y comunitario, tendentes a garantizar en mayor o menor medida la transparencia algorítmica y el derecho a una explicación, esto es, el derecho a conocer cómo funciona el concreto proceso de toma de decisión con indicación de los datos específicos de la persona en base a los cuales se ha tomado una decisión o se ha hecho un perfilado[75]. El objetivo de todas ellas es eliminar, o al menos mitigar, la opacidad tanto en el *uso* como en el *contenido* de la información sobre el algoritmo y se asientan sobre una serie de principios como el de transparencia desde el diseño, minimización de sesgos y rendición de cuentas[76].

---

[75] Más allá del RGPD, a nivel europeo contamos con algunas normas que regulan y exigen transparencia algorítmica en ámbitos concretos. Por ejemplo, respecto a las plataformas que prestan servicios de intermediación en línea, el *Reglamento (UE) 2019/1150 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, sobre el fomento de la equidad y la transparencia para los usuarios profesionales de servicios de intermediación en línea* (conocido como Reglamento P2B), que obliga a plataformas y motores de búsqueda a dar publicidad a los parámetros principales de funcionamiento de sus algoritmos de prelación de ofertas (arts. 5 y 7). Por otro lado, la *Directiva (UE) 2019/2161 del Parlamento Europeo y del Consejo de 27 de noviembre de 2019 por la que se modifica la Directiva 93/13/CEE del Consejo y las Directivas 98/6/CE, 2005/29/CE y 2011/83/UE del Parlamento Europeo y del Consejo*, en lo que atañe a la mejora de la aplicación y modernización de las normas de protección de los consumidores de la Unión, establece que los comerciantes pueden personalizar el precio de sus ofertas para determinados consumidores o determinadas categorías de consumidores basándose en la toma de decisiones automatizada y la elaboración de perfiles a partir de su comportamiento, pero en tales casos se les deberá informar acerca de los principales parámetros por defecto que determinan la clasificación de las ofertas mostradas y su importancia relativa frente a otros parámetros, a fin de que puedan tener en cuenta los riesgos potenciales de su decisión de compra. También en muchos instrumentos de *soft law* provenientes de la UE se aboga por la transparencia, información y explicabilidad de los sistemas algorítmicos. Vid., por ejemplo, las *Resoluciones del Parlamento Europeo de 12 de febrero de 2020 sobre los procesos automatizados de toma de decisiones*, o la de *20 de octubre de 2020 sobre aspectos éticos de la IA*. Y también contamos en el ámbito europeo con un reconocimiento de derechos digitales del que se desprende el derecho a conocer que estamos siendo objeto decisiones algorítmicas. Se trata de la *Declaración sobre los Derechos y Principios Digitales*, de 23 de enero de 2023, donde se recoge el compromiso de «velar por un nivel adecuado de transparencia en el uso de los algoritmos y la inteligencia artificial y porque las personas estén informadas y capacitadas para utilizarlos cuando interactúen con ellos».

[76] Particularmente relevante en el tema que nos ocupa es la *Ley 15/2022 integral de igualdad,* que en su art. 23, establece el deber de las administraciones públicas de poner en marcha mecanismos para que los algoritmos involucrados en la toma de decisiones tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, y que se priorice la transparencia en el diseño y la implementación, así como la capacidad de interpretación de las decisiones adoptadas por los mismos, precepto que, no obstante su relevancia a la hora de sentar parámetros generales, ha sido considerado por la doctrina como una mera declaración de intenciones que deberá concretarse en medidas específicas a cumplir por las administraciones y las empresas. En el ámbito público, algunas leyes de transparencia autonómicas ya recogen deberes de información en relación con el uso de algoritmos. Así, la pionera *Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunidad Valenciana*, en su art. 16 (letra l). Por otro lado, y aunque no tiene carácter vinculante, la *Carta de Derechos Digitales* española reconoce a los ciudadanos una serie de derechos frente a la Administración cuando se utilicen sistemas inteligentes en la toma de decisiones, que se traducen en cargas o deberes de información y transparencia para los entes públicos *(vid.,* ap.XVIII.6). En el

Además, contamos con pronunciamientos judiciales y de órganos administrativos que se posicionan a favor de garantizar un nivel de transparencia algorítmica mucho mayor y obligaciones de información más exigentes en los casos en que los sistemas de IA para la automatización de decisiones se utilicen en el ámbito público[77]. Todo ello contribuye, sin duda alguna, a luchar contra algunos de los principales obstáculos que plantea la discriminación algorítmica como son su detección y prueba.

Asimismo, todas las soluciones propuestas o aprobadas pretenden garantizar la transparencia algorítmica sin afectar a los derechos de autor y los secretos empresariales. De hecho, encontrar un equilibrio entre la propiedad intelectual de ciertos métodos de procesamiento y la necesidad de garantizar la transparencia, imparcialidad y equidad algorítmica, ha sido considerado por las instituciones europeas como un objetivo imprescindible desde que comenzaron a trabajar sobre la IA hasta sus más recientes y relevantes propuestas normativas[78].

De todas las medidas legales para hacer frente a la opacidad de los sistemas de IA, en el presente trabajo nos centraremos en las contenidas en el RGPD y en el RIA. Por lo que se refiere al RIA, su reciente aprobación y la importancia de las medidas que contiene respecto a la transparencia algorítmica y la explicabilidad, justifican sobradamente que le dediquemos una atención

---

Derecho laboral español, junto con la conocida como *Ley Rider* (a la que haremos referencia posteriormente), otra norma con la que se trata de hacer frente a la opacidad algorítmica y fomentar la transparencia de la IA es la *Ley 3/2023, de 28 de febrero, de Empleo*. Dicha norma establece la creación de herramientas de apoyo a la toma de decisiones basadas en evidencias estadísticas para la mejora de la empleabilidad de las personas. En concreto, es el art. 17 el que se ocupa de la «toma de decisiones fundamentada en el análisis de datos, las evidencias estadísticas y el análisis del mercado de trabajo».

[77] Además de la sentencia de 5 de febrero de 2020 del Tribunal de lo Civil de la Haya dictada en el caso del algoritmo Syri, también se obligó a dar acceso al código fuente de diferentes sistemas automatizados empleados por el sector público en los siguientes asuntos: Tribunal Regional Administrativo del Lazio-Roma, Sección III bis, Sentencias Nº 3769, de 22 de marzo de 2017 y Nº 10964, de 13 de septiembre de 2019; Comisión Francesa de Acceso a los Documentos Administrativos, Decisiones Nº 20144578, de 8 de enero de 2015 y Nº 20180276, de 19 de abril de 2018. En nuestro país, la Resolución de 21 de septiembre de 2016 de la Comisión de Garantía del Derecho de Acceso a la Información Pública (GAIP), que estima las Reclamaciones 123/2016 i 124/2016 (acumuladas) sobre el algoritmo matemático que determina la selección de los miembros de los tribunales correctores de las pruebas de acceso a la Universidad (PAU) y DNI de los candidatos presentados, y su Resolución 200/2017, de 21 de junio relacionada con las anteriores. Por lo que se refiere al CTBG, las resoluciones R/0058/2021 de 20 de mayo, sobre el acceso al algoritmo para calcular pensiones de la TGSS; RT/253/2021 de 19 de noviembre y RT/ 748/2021 de 10 de enero 2022 sobre acceso al código fuente de una aplicación utilizada para el sorteo de tribunales asociados a procesos selectivos para educación en Madrid, y la Resolución del Expediente 551-2023 sobre el acceso a información del Sistema de Seguimiento Integral de los casos de Violencia de Género (Sistema VioGén).

[78] Vid., por ejemplo, el art. 78 RIA; el Considerando 20 y el art. 3.4 PDRC; y los Considerandos 31 y 32 y art. 8 apartados 2,3 y 4 PDPD.

especial. Por lo que respecta al RGPD, la novedosa jurisprudencia sentada por el TJUE sobre su art. 22, precepto básico y fundamental para hacer frente a los riesgos del *machine learning* en la toma de decisiones basadas en datos personales, hace que sea necesario analizar con particular detenimiento dicho precepto y su interacción con el nuevo art. 86 RIA.

4.2 La transparencia y la explicabilidad en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, de protección de datos (RGPD): especial referencia a las decisiones automatizadas del art. 22

Los principios sobre los que se asienta el RGPD[79], recogidos en su art. 5 y entre los que se encuentra el de transparencia, obligan a responsables y encargados del tratamiento a comunicar al interesado cómo y por qué se recogen y utilizan sus datos personales, así como a garantizar su exactitud, entre otras cosas. En virtud del principio de transparencia, desarrollado en los Considerandos 39 y 58, debe quedar totalmente claro para las personas físicas que se están recogiendo, utilizando, consultando o tratando de otra manera sus datos personales, así como la medida en que dichos datos son o serán tratados. Además, toda la información y comunicación relativa al tratamiento debe ser fácilmente accesible y fácil de entender, utilizando un lenguaje sencillo y claro y, en su caso, visualizándose[80].

Más allá de las exigencias derivadas del citado principio, el RGPD establece una serie de garantías y derechos para hacer frente a la opacidad algorítmica cuando el interesado es objeto de una decisión plenamente automatizada basada en sus datos personales y con efectos jurídicos o similares en él: por un lado, los arts.13.2 f), 14.2 g) y 15.1.h) obligan a informarle sobre la *existencia de decisiones automatizadas,* incluida la elaboración de perfiles, y al menos en tales casos, a darle *información significativa sobre la lógica aplicada, así como la importancia y las consecuencias que dicho tratamiento puede tener para su persona.* Por tanto, se consagra un derecho de información reforzado o ampliado exigible cuando concurren los presupuestos mencionados; por otro lado, el art. 22.3 establece que, en tales casos, el responsable del tratamiento adoptará las

---

[79] Dichos principios son: licitud, lealtad y transparencia (art. 5.1.a); limitación de la finalidad (art.5.1.b); minimización de datos (art. 5.1.c); exactitud (art. 5.1.d); limitación del plazo de conservación (art. 5,1.e); integridad y confidencialidad (art. 5.1.f); y responsabilidad proactiva (art. 5.2).

[80] Se destaca también en el Considerando 58 que la transparencia «es especialmente pertinente en situaciones en las que la proliferación de agentes y la complejidad tecnológica de la práctica hagan que sea difícil para el interesado saber y comprender si se están recogiendo, por quién y con qué finalidad, datos personales que le conciernen, como es en el caso de la publicidad en línea».

medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, y, como mínimo, el *derecho a obtener intervención humana* por parte del responsable, a *expresar su punto de vista* y a *impugnar la decisión* que le afecta.

No hay unanimidad acerca de si el RGPD consagra un «derecho de explicabilidad algorítmica» o «derecho a abrir la caja negra de los algoritmos». Mientras el derecho de información *ex ante* está adecuadamente regulado en el RGPD -como se deriva de los arts. 5.1, 13, 14 o 15 que acabamos de mencionar- y parece claro que opera con anterioridad a que el sistema adopte la decisión o haga el perfilado, se discute su existencia *ex post*, esto es, después de elaborada la decisión o el perfil de la persona. La controversia se basa, principalmente, en que dicho derecho no se reconoce en el texto articulado sino simplemente en uno de sus considerandos, en concreto en el número 71, según el cual, el afectado por la decisión automatizada goza del «derecho a obtener intervención humana, a expresar su punto de vista, *a recibir una explicación de la decisión tomada después de tal evaluación* y a impugnar la decisión»[81]. Teniendo en cuenta que los considerandos no son en sí mismos normas, si la concreta decisión o el perfilado adoptados por el sistema perjudican al interesado, no hay consenso acerca de si éste tendría derecho a que se le informara sobre la *lógica decisional* (distinta de la *lógica aplicada* a la que se refieren los art. 13,14 y 15[82]), es decir, a que se le explicara el proceso decisorio que ha

---

[81] Sobre esta controversia doctrinal, puede verse LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.*, pp. 197 y ss, y doctrina citada en notas 718 y 719. También sobre el asunto, NÚÑEZ SEOANE, J., «El derecho a la información y acceso al funcionamiento de los algoritmos que tratan datos personales», en *La regulación de los algoritmos*, coord. por Gustavo Manuel Díaz González y Alejandro José Huergo Lora (dir.) 2020, pp.307-308.

[82] Diferencia con claridad entre lo que sería la lógica aplicada y la lógica decisional, GIL MEMBRADO, C., «Daños producidos por la IA: la opacidad del algoritmo y el efecto de caja negra», en N. Álvarez Lata/J.M. Busto Lago, *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*, 2024, p. 521. La primera estaría reconocida en los arts. 13 y 14 RGPD y obligaría a informar sobre el sistema de IA en general, mientras que la segunda obligaría a abordar qué proceso decisorio ha conducido a un determinado resultado de salida. En su opinión, el RGPD sólo reconoce la primera (esto es, el derecho a conocer la lógica aplicable -*ex ante*-, pero no el derecho a conocer la explicabilidad -*ex post*-). Para MEDINA GUERRERO, M., «El derecho a conocer los algoritmos utilizados en la toma de decisiones. Aproximación desde la perspectiva del derecho fundamental a la protección de datos personales», *Teoría y realidad constitucional*, nº 49, 2022, p. 161, el deber de proporcionar «información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas» previsto en los arts. 13.2.f) y 14.2.g) RGPD, sólo puede traducirse en la práctica en una explicación *ex ante* sobre la *funcionalidad del sistema* en la medida en que debe llevarse a cabo en el momento en que se obtengan los datos personales (ex art. 13.2) o en un plazo muy breve (ex art. 14.3.a). Al realizarse la notificación antes de que el sistema decisional haya adoptado la solución concreta, la información que debe dar el responsable no puede relacionarse con el resultado de ninguna decisión específica, sino que deberá centrarse en el sistema considerado como un todo abstracto. Por lo que, como es lógico, en esta fase la obligación de información no puede traducirse en la exigencia de asegurar la explicabilidad o la fundamentación de una decisión en particular. Y en sentido similar,

conducido a que el sistema arrojara ese resultado concreto, así como si este derecho incluiría, a su vez, el derecho a acceder a los datos de entrenamiento del sistema.

La cuestión no es baladí si se tiene en cuenta que la solicitud de un interesado para investigar posibles errores o sesgos discriminatorios de la IA se producirá siempre *ex post*, es decir, una vez adoptada la decisión discriminatoria[83]. Para un sector doctrinal, al que nos adherimos, se podría defender la existencia de este derecho a la explicabilidad *en ambos momentos*. Una interpretación sistemática de los derechos de información de los arts. 13, 14 y 15, junto con los derechos del art. 22.3 y el Considerando 71, implicaría reconocer que el derecho a la explicación está incorporado siquiera implícitamente en el art. 22.3[84]. Así lo mantiene entre nuestros autores NÚÑEZ Seoane para quien no cabe duda de que se reconoce este derecho en ambas sedes teniendo en cuenta que es el resultado o consecuencia del ejercicio por el interesado de los previos derechos o garantías a la intervención humana y a formular las alegaciones que prevé el artículo 22.3; así, una vez solicitada la intervención humana y formuladas las alegaciones pertinentes, el responsable deberá ofrecer «una explicación de la decisión tomada después de tal evaluación» a la que se refiere el Considerando 71, ya que solo después de que sus alegaciones hayan

---

apunta PALMA ORTIGOSA, A., «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», *Revista General de Derecho Administrativo*, n. 50, 2019, p. 14/25, que «el derecho a la explicación exige del responsable una aportación de información mucho más personalizada sobre la decisión algorítmica que aquella que debía aportar en un primer momento el responsable conforme a los Arts. 13.2.f), 14.2.g), 15.1.h) RGPD, información previa que no tenía todavía en cuenta la decisión automatizada del particular basada en los datos personales aportados por este último».

[83] LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.*, p. 197.

[84] En este sentido, SELBST, A. D./POWLES, J., «Meaningful information and the right to explanation», *International Data Privacy Law*, Vol. 7, n.º 4, 2017, pp.233 y ss, GOODMAN., B/FLAXMAN, S., «European Union regulations on algorithmic decision-making and a "right to explanation"» presented at ICML Workshop on Human Interpretability in Machine Learning, New York, 2016, https://arxiv.org/abs/1606.08813v3; BIBAL A./LOGNOUL, M./DE STREEL, A./FRÉNAY, B., «Legal requirements on explainability in machine learning», *Artificial Intelligence & Law*, 29(2), 2021, pp. 149-169. También entre nuestros autores, NAVAS NAVARRO, S., *Daños ocasionados …, cit.,* p. 72, NÚÑEZ SEOANE, J., «El derecho a la información y acceso al funcionamiento de los algoritmos que tratan datos personales», en *La regulación de los algoritmos*, coord. por Díaz González y Huergo Lora (dir.) 2020, LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.loc.cit.* En sentido contrario, la falta de refrendo legal y el carácter no vinculante de los considerandos, ha llevado a otro sector doctrinal a entender que en el RGDP no existe un derecho de explicabilidad. Manteniendo esta última postura, WACHTER, S., MITTELSTADT, B., y FLORIDI, L., «Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation», *International Data Privacy Law*, Vol. 7, n. 2, 2017. Estos autores afirman que la omisión del derecho a la explicación en el artículo 22.3 es intencionada, por lo que niegan su reconocimiento. También, en nuestra doctrina, GIL MEMBRADO, C., *op.cit.loc.cit.*, considera que el RGPD no reconoce el derecho a la explicabilidad *ex post.*

sido revisadas, evaluadas y explicadas por una persona, podrá ejercer en condiciones su derecho de impugnación[85].

Junto con las vacilaciones acerca de la propia *existencia* del derecho a la explicabilidad, también se ha discutido (y criticado) su *alcance* y *contenido*. Determinar cuándo opera este derecho obliga a analizar con más detalle el art. 22 RGPD por cuanto los derechos y garantías mencionadas se aplican sólo en los casos de decisiones individuales automatizadas y perfilados a que éste se refiere.

4.2.1 Ámbito de aplicación del art. 22 RGPD

El art. 22 RGPD es el punto de enlace entre la IA y la protección de datos, puesto que las decisiones automatizadas y la elaboración de perfiles son dos de sus aplicaciones más frecuentes[86]. Se trata de uno de los pocos preceptos que prevé una serie de garantías jurídicas en los casos en que se adopten decisiones automatizadas, ya sea mediante sistemas de IA o no. Por tanto, si el tratamiento de datos personales, la elaboración de perfiles o la toma de decisiones sobre una persona física es realizada por un componente IA, tendrá que someterse al RGPD. En caso contrario, esto es, si no se tratan datos personales, no será necesario. El problema reside en que en muchos casos no es sencillo determinar si durante una etapa del ciclo de vida de un sistema basado en IA se tratan o no datos personales[87].

Tanto el ámbito de aplicación del art. 22, como los derechos y garantías que implica y su concreta efectividad cuando la automatización de decisiones se lleva a cabo a través de sistemas de IA, han sido objeto de críticas y comen-

---

[85] NÚÑEZ SEOANE, J., *op.cit.*, p. 308

[86] HERRERA DE LAS HERAS, R., «Protección de datos e inteligencia artificial», en Cruz Blanca/Lledó Benito (coords.), *La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0*, 2021, p. 654. No obstante, como matiza REBOLLO DELGADO, L. *Inteligencia artificial y derechos fundamentales*, Madrid, 2023, p. 107, estas dos funcionalidades hoy en día ya no son troncales de la IA y la aplicación del precepto únicamente a ellas se manifiesta claramente insuficiente teniendo en cuenta la amplitud de posibilidades que ofrece la IA. Para una explicación detallada sobre las diferencias entre estas dos figuras jurídicas recogidas en el RGPD que guardan relación con los procesos de tratamiento de datos y que dan lugar a inferencias o predicciones sobre comportamientos o preferencias de las personas, puede verse LAZCOZ MORATINOS, G., *op.cit.,* p. 288. Baste mencionar aquí que las decisiones automatizadas tienen un ámbito de aplicación que puede solaparse parcialmente con la elaboración de perfiles o derivar de ésta. En ocasiones, la toma de decisiones automatizada requiere previamente la elaboración del perfil de la persona intentando hacer una predicción sobre sus características o comportamiento; en otros casos, la decisión se toma sobre la base de datos referentes a la actividad de la persona sin necesidad de elaborar un perfil.

[87] Así lo afirma la AEPD, *Adecuación*…, cit., p.14.

tarios por la doctrina[88]. Además, en fechas muy recientes, el TJUE se ha pronunciado por primera vez sobre el ámbito de aplicación de dicho precepto, en su sentencia de 7 de diciembre de 2023, asunto C-634/21 (*OQ y Land Hessen, con intervención de SCHUFA Holding AG*), sentando doctrina importante al respecto. Es conveniente, por ello, detenerse en el análisis de tales cuestiones.

La primera garantía que recoge el art. 22 RGPD consiste, precisamente, en *prohibir con carácter general que las personas puedan ser objeto de decisiones basadas únicamente en el tratamiento automatizado de sus datos personales*, incluyendo la elaboración de perfiles, *siempre que dichas decisiones produzcan efectos jurídicos sobre ellas o les afecten significativamente de modo similar* (art. 22.1). No obstante, se establecen una serie de casos en los que excepcionalmente se permiten dichas decisiones y que restan eficacia a la prohibición, a saber: a) que sean necesarias para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento; b) que estén autorizadas por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado; y c) que el interesado haya dado su consentimiento explícito. Analizaremos a continuación tanto las excepciones como la regla general contenida en el precepto.

A) Excepciones

La primera excepción plantea muchas dudas interpretativas acerca de cuándo «es necesaria» una decisión de este tipo para la celebración o ejecución de un contrato, ya que lo normal es que el uso de sistemas para automatizar decisiones responda a razones de conveniencia de una de las partes, más que de necesidad[89]. Además, la «necesidad» debe interpretarse de manera estricta, lo que significa que el tratamiento automatizado debe ser esencial para poder perfeccionar el acuerdo o cumplir con las obligaciones contractuales. Así, esta base de legitimación podrá alegarse en sectores u organiza-

---

[88] Por todos, PALMA ORTIGOSA, A., *op.cit.,* y REBOLLO DELGADO, L., *op.cit.,* pp. 101 y ss. También afirma con rotundidad OTTOLIA, A., *Derecho, Biga Bata e Inteligencia Artificial*, Valencia, 2018, p.102, que «[…] la aplicación real del precepto es decepcionante, ya que resulta debilitado por las excepciones que prevé».

[89] HUERGO LORA, A., «Una aproximación a los algoritmos desde el derecho administrativo», en A. HUERGO LORA (dir.) y G. DÍAZ GONZÁLEZ (coord.), *La regulación de los algoritmos*, Navarra, 2020, p. 62. De hecho, el GT 29 reconoce en sus *Directrices sobre decisiones individuales automatizadas*, cit., p. 26, que los responsables del tratamiento pueden, con fines contractuales, acudir a los procesos de decisiones basadas únicamente en el tratamiento automatizado de datos por considerar que es la mejor manera de conseguir un determinado objetivo, ya que «la participación humana rutinaria puede resultar poco práctica o imposible debido a la magnitud de los datos tratados».

ciones donde es habitual realizar una evaluación previa del interesado que pretende formalizar un contrato con el responsable. Piénsese por ejemplo en sectores como el bancario, el de seguros o el laboral. De hecho, en el caso de la concesión de créditos es obligatorio realizar una evaluación previa de la solvencia patrimonial del solicitante[90]. Para ello, el responsable, antes de concertar dicho contrato, valorará al interesado con el objetivo de analizar la viabilidad futura del acuerdo. En definitiva, lo realmente importante para que se pueda aplicar esta base legitimadora del tratamiento es que el responsable demuestre que no hay otro método igualmente efectivo y menos invasivo para la intimidad con el que se pueda lograr el mismo objetivo. Si lo hubiera, dicho tratamiento no podría calificarse de «necesario»[91].

Respecto a la segunda excepción, parece que se permiten las decisiones plenamente automatizadas cuando estén autorizadas legalmente, es decir, se permite que el Derecho de la UE o de los Estados miembros pueda establecer normas donde se autorice a los responsables a llevar a cabo este tratamiento automatizado de datos personales para la toma de decisiones. Del propio tenor de esta disposición se desprende que el Derecho nacional que autorice la adopción de una decisión individual automatizada debe establecer medidas adecuadas para salvaguardar los derechos, libertades e intereses legítimos del interesado[92]. Además, el tratamiento deberá respetar los requisitos establecidos en los arts. 5 y 6 RGPD[93].

Respecto a la tercera excepción, el consentimiento del interesado como base legitimadora para el tratamiento de datos personales es la clave de bóveda de la normativa de protección de datos. Sin embargo, no garantiza en absoluto un conocimiento y comprensión suficientes de las consecuencias

---

[90] Vid., art. 18 de la Orden EHA/2899/2011, de 28 de octubre, de transparencia y protección del cliente de servicios bancarios y art 29 de la Ley 2/2011, de 4 de marzo, de Economía Sostenible. Sobre la obligación de los bancos de evaluar la solvencia del cliente antes de la concesión de un crédito, puede verse MARÍN LÓPEZ, M.J., «La obligación de evaluar la solvencia del prestatario en la Ley 5/2019, de contratos de crédito inmobiliario», *Revista de Derecho Patrimonial*, n. 50/2019, accesible en https://manueljesusmarin.es/wp-content/uploads/2022/02/doctrin_084.pdf.

[91] Así lo explica el GT29 en sus *Directrices sobre decisiones individuales automatizadas*, cit., p.26.

[92] El Considerando 71 RGPD establece que las medidas a adoptar deben incluir, en particular, la obligación del responsable del tratamiento de utilizar procedimientos matemáticos o estadísticos adecuados, de aplicar las medidas técnicas y organizativas apropiadas para garantizar que se reduzca al máximo el riesgo de error y se corrijan errores, y de asegurar los datos personales de forma que se tengan en cuenta los posibles riesgos para los intereses y derechos del interesado y se impidan, entre otras cosas, los efectos discriminatorios en las personas físicas.

[93] Como indica el TJUE, en su sentencia de 7 de diciembre de 2023, párrafos 65, 66 y 68, los Estados miembros no podrán adoptar normativas que autoricen las decisiones automatizadas o elaboración de perfiles incumpliendo los requisitos establecidos en los arts. 22.2.b) y 22.4 y en los arts. 5 y 6 RGPD. Para un mayor desarrollo de esta excepción puede verse PALMA ORTIGOSA, A., *op.cit.*, p. 9/25.

que implica la cesión y el tratamiento de dichos datos. Lo mismo ocurre con las decisiones automatizadas puesto que los interesados no son conscientes de la trascendencia e impacto de sus resultados ni de los posibles sesgos discriminatorios que pueden esconder. En cualquier caso, este consentimiento debe cumplir con los requisitos del art. 4.11 en relación con el Considerando 42 RGPD, es decir, debe ser informado, específico, inequívoco, libre y revocable[94].

Pues bien, en tales supuestos excepcionales, y más concretamente en los previstos en las letras a) y c), el precepto recoge una serie de garantías «mínimas» que deberá adoptar el responsable para salvaguardar los derechos e intereses legítimos del afectado por la decisión automatizada: concretamente, el *derecho a obtener intervención humana* por parte del responsable, a *expresar su punto de vista* y *a impugnar*[95] la decisión (art.22.3). A las que se debe añadir, tal como menciona el Considerando 71, la *información específica* al interesado. Las mismas garantías, al menos, deberá adoptar el Estado que autorice este tipo de decisiones en su ordenamiento en virtud del art. 22.2.b).

A continuación, se establecen dos importantes precisiones respecto a este tipo de decisiones que deben tenerse en consideración: por un lado, que la decisión automatizada no puede basarse en las categorías de datos especialmente protegidos por el art. 9.1 (art. 22.4); y por otro lado, que mediante Ley se podrá limitar el alcance de las obligaciones y derechos establecidos en el art. 22 en los numerosos supuestos del artículo 23 (art.23.1).

Respecto al tratamiento de las categorías especiales de datos (también conocidos como datos sensibles) vienen a coincidir prácticamente con las denominadas *categorías sospechosas* o *motivos protegidos* de la normativa antidiscriminatoria. Se trata, concretamente, de datos que revelen el origen étnico o racial, las opiniones políticas, religiosas o filosóficas, los datos genéticos o biomédicos y aquellos otros referentes a la vida u orientación sexual de la persona[96]. Por tanto, cualquier sesgo en los resultados del sistema causado al tener en cuenta este tipo de datos de manera voluntaria o por error, supon-

---

[94] Analiza cada uno de estos requisitos, CASADESÚS RIPOLL, P., *La responsabilidad civil por el uso discriminatorio de los datos personales a través de la inteligencia artificia*, Granada, 2022, pp. 149-154.

[95] Sobre este concreto derecho, RIVAS VALLEJO, P. «Análisis…», *cit.*, p. 460 dice que «Tampoco se matiza si el derecho a impugnar la decisión tiene un marco de acción reducido a la interacción con "el responsable del tratamiento" o si tiene un contenido más amplio equivalente al derecho a reclamar contra la decisión empresarial, que debería entenderse en sentido interno, pues la legislación garantiza en todo caso el derecho a la reclamación administrativa o judicial (art. 24 CE)».

[96] Salvo en determinadas circunstancias mencionadas en el ap. 2, dispone el art. 9.1 que «*Quedan prohibidos el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométri-*

dría el incumplimiento de este artículo cuando no concurra causa legítima para su tratamiento, además del incumplimiento de la prohibición legal de no discriminar.

Debe tenerse en cuenta que la prohibición de tratar categorías especiales de datos tiene excepciones. Así, cuando concurran una serie de circunstancias recogidas en el art. 9.2 su tratamiento estará permitido bajo ciertas condiciones. En concreto, se autoriza el tratamiento de datos sensibles *cuando sea necesario por razones de interés público esencial,* sobre la base del Derecho de la Unión o de los Estados miembros (letra g). Precisamente dicho fundamento ha sido el esgrimido en el RIA para justificar el tratamiento de estas categorías especiales de datos a través de sistemas IA. Evitar la discriminación que podría provocar el sesgo de los sistemas de IA se considera una cuestión de orden público esencial en virtud de la cual los proveedores deben ser capaces de tratar ese tipo de datos «*con carácter excepcional, en la medida en que sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo, con sujeción a las garantías adecuadas para los derechos y libertades fundamentales de las personas físicas*[97] *y tras la aplicación de todas las condiciones aplicables establecidas en el presente Reglamento, además de las condiciones establecidas en los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y la Directiva (UE) 2016/680*»[98]. Es el art. 10.5 el que se encarga de establecer las condiciones para que se pueda llevar a cabo dicho tratamiento.

B) La prohibición general del art. 22 RGPD: presupuestos de aplicación a la luz de la última jurisprudencia del TJUE en su sentencia de 7 de diciembre de 2023

Dos son los presupuestos que exige el art. 22.1 para que opere la prohibición que contiene: por un lado, que se trate de decisiones basadas *únicamente* en el tratamiento automatizado de datos personales; y por otro lado, que dichas decisiones tengan *efectos jurídicos o significativamente similares* en el interesado.

---

cos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física».

[97] En la Orientación General del Consejo sobre el RIA, de 25 de noviembre de 2022, el texto del art. 10.5 mencionaba como salvaguardias adecuadas para los derechos y libertades fundamentales, el establecimiento de «*limitaciones técnicas a la reutilización y la utilización de las medidas de seguridad y protección de la privacidad más recientes, tales como la seudonimización o el cifrado, cuando la anonimización pueda afectar significativamente al objetivo perseguido*».

[98] Vid., Considerando 70.

a) Primer presupuesto: que se trate de una decisión exclusivamente automatizada

El primer presupuesto implica el uso de sistemas que de forma automática y sin intervención humana significativa adopten una decisión que afecte a la persona (por ejemplo, decidir qué candidatos para un puesto de trabajo se descartan en un proceso de selección y quiénes pasan a la fase de entrevista, rechazar automáticamente la concesión de un crédito o la concertación de un seguro, etc). Como se encargó de aclarar el GT29, para que se trate de decisiones exclusivamente automatizadas no debe haber una intervención humana responsable que analice la información procesada y cuente con poder o capacidad de modificar la decisión[99].

Surge la duda de si la prohibición se aplicaría a los casos en que esa intervención humana relevante se limitara a una o varias partes del proceso de decisión, pero no a todas. Podría defenderse que, en la medida en que dicha intervención humana significativa exista (aunque sea en una etapa del proceso), las decisiones no quedarían bajo el ámbito de aplicación del art. 22. Por tanto, en tales casos el afectado no gozaría de los derechos y garantías anteriormente comentadas. No obstante, creemos más acertado mantener que para poder eludir la aplicación del precepto, la existencia de intervención humana debería apreciarse en cada una de las decisiones parceladas dirigidas al mismo fin y no solamente en alguna parte del proceso. Así, por ejemplo, si un algoritmo selecciona las 10 mejores candidatas a un puesto de trabajo de entre las 1000 que se postulan, descartando automáticamente a las demás, existiría una decisión automatizada respecto de la que habría que garantizar los derechos recocidos en la norma, aunque las 10 últimas fueran evaluadas por una persona[100]. Por el contrario, el art. 22 sería inaplicable si en ese proceso ha intervenido un supervisor que ha decidido quién pasa a la segunda fase y quién no, sobre la base de la recomendación que hace el sistema IA de apoyo a la decisión humana (y que, si lo considera oportuno, puede no seguir),

---

[99] *Directrices sobre decisiones individuales automatizadas*, cit., p. 23. Sobre lo que implica esta exigencia puede verse también RIVAS VALLEJO, P. «Análisis…», *cit.,* p. 457 y ss.

[100] MINISTERIO DE TRABAJO Y ECONOMÍA SOCIAL, *Información algorítmica en el ámbito laboral. Guía Práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral,* mayo 2022, p.10. Accesible en file:///C:/Users/ASUS/Downloads/algortca_laboral-1.pdf. Como explica RIVAS VALLEJO, P. «Análisis…», *cit.,* p. 456, «la norma parece apuntar a un ámbito restringido del citado derecho a la explicabilidad, no extensible, por tanto, a decisiones automatizadas en parte del proceso de decisión, aun cuando esta parte sea especialmente relevante en el conjunto de factores que conducen a la decisión final». El problema de mantener esta interpretación, como sigue diciendo la autora, es que «si se considera que han sido presentados mil currículos, de los cuales han sido preseleccionados quince para su examen humano, el grueso del sesgo, donde probablemente radique la mayoría de discriminaciones por diversas causas protegidas, habrá quedado incluido en una fase íntegramente automatizada, sobre la cual no se ofrecerá explicación alguna».

salvo que esa intervención humana hubiera sido exclusivamente simbólica o ficticia, asumiendo sin más lo decidido por el sistema[101].

En definitiva, para eludir tanto la prohibición como las obligaciones que impone el art. 22 a los responsables del tratamiento cuando hay base legitimadora del tratamiento automatizado, debe quedar claro que la intervención humana debe ser significativa en el sentido de ser realizada por una persona con competencia y autoridad sobre la decisión y que valore toda la información disponible, no siendo suficiente con que se limite a convalidar, replicar o reproducir acríticamente la decisión o recomendación del sistema. Su intervención no debe consistir en una mera supervisión superficial de los resultados del algoritmo sino en una participación real, relevante o sustancial en el proceso decisorio.

Del análisis de este primer presupuesto puede extraerse, como conclusión, que la regulación no afecta a todas las predicciones algorítmicas sino solo a aquellas que se integran en sistemas de decisión plenamente automatizados[102]. Por ello, y sobre la base de que en la práctica lo normal es que las decisiones no suelan tomarse de forma plena o puramente automática (sobre todo cuando son de una entidad considerable), sino como un elemento más de juicio, junto con otros, para tomar la decisión final, el art. 22 RGPD perdería cierta relevancia puesto que se aplica exclusivamente a una de las distintas formas de utilización de estos sistemas algorítmicos[103]. Es decir, la prohibición que contiene gozaría de escasa aplicación en tanto en cuanto no abarca las decisiones semiautomatizadas y, en la actualidad, no es tan frecuente la toma de decisiones plenamente automatizadas que cumplan las condiciones que exige la norma. Lo más normal es que se utilicen los resultados de estos sistemas como elemento auxiliar que el operador humano puede tener en cuenta, o no, para adoptar su decisión final[104].

Dicha limitación supone una importante traba en la lucha contra la opacidad de los sistemas de IA (y por lo que a nosotros interesa, contra la discriminación algorítmica) por cuanto los sistemas que se usan como apoyo a la decisión también pueden arrojar resultados incorrectos y discriminatorios al

---

[101] Como mantiene EGUÍLUZ CASTAÑEIRA, J.U., *op.cit.,* p. 344, si no se entendiera que esa intervención humana no ha de limitarse a reproducir la decisión del sistema, las empresas podrían eludir fácilmente la aplicación del art. 22, destinando un empleado a supervisar decisiones automatizadas, que en la práctica se limitara a asumir acríticamente sus resultados.

[102] En este sentido, por todos, HUERGO LORA, A., «Una aproximación…», cit., p. 60 o GINÈS i FABRELLAS, A., *op.cit.,* p. 319. También se hace referencia a esta idea en MINISTERIO DE TRABAJO Y ECONOMÍA SOCIAL*, Información algorítmica en el ámbito laboral…, cit.,* p. 15, cuando se dice que la regulación europea usa un concepto restrictivo de algoritmo en la medida en que exige que tome decisiones íntegramente automatizadas.

[103] PALMA ORTIGOSA, A., *op.cit.,* p. 5/25.

[104] Por todos, HUERGO LORA, A., «Una aproximación…», cit., p. 71.

contener, por ejemplo, sesgos en el diseño o en los conjuntos de datos utilizados[105], y principalmente, por el denominado «sesgo de la automatización», es decir, la tendencia comprobada de que los humanos confiamos o creemos (erróneamente) que el resultado ofrecido por la máquina es mucho más acertado, objetivo y neutral que el nuestro, cuando ambos juicios no coinciden[106].

Por ello, en aras de una mejor defensa de los intereses del perjudicado por la decisión, señala el GT29 que, aunque la decisión automatizada o la elaboración de perfiles no cumplan la definición del art. 22.1, seguiría siendo aconsejable ofrecer la información ampliada o reforzada prevista en el RGPD[107]. Ahora bien, a nuestro modo de ver, más allá de esta recomendación o consejo, lo que habría que plantearse sería la expresa ampliación del ámbito de aplicación de este derecho de información específico y demás garantías previstas en el precepto a los casos de decisiones semiautomatizadas, tal y como por ejemplo ha hecho nuestro legislador laboral en el art. 64.4 ET[108] y como parece sostener también el TJUE en su sentencia de 7 de diciembre de 2023 que será analizada *infra*.

b) Segundo presupuesto: efectos jurídicos o similares

Respecto al segundo presupuesto exigido por el art. 22 RGPD, esto es, que la decisión automatizada produzca *efectos jurídicos o significativamente simi-*

---

[105] GINÈS i FABRELLAS, A., *op.cit.,* p. 319

[106] En SOLAR CAYÓN, J.I., *op.cit.,* p. 170, se apunta que la literatura científica en sociología cognitiva y economía conductual muestra que es psicológicamente muy difícil desatender las recomendaciones de los sistemas algorítmicos, por lo que lo más probable es que las personas acaben siguiéndolas, especialmente, cuando se promueve el uso de estos sistemas bajo el argumento y con el objetivo de contribuir a erradicar la subjetividad y los errores humanos. Y BELTRÁN DE HEREDIA, I., «Automatización y obsolescencia humana», en *Retos jurídicos de la inteligencia artificial*, coord. por Agustí Cerrillo i Martínez y Miquel Peguera Poch, 2020, p. 120, manifiesta que se trata de un tipo de sesgo que, a pesar de estar presente en muchos ámbitos, supone un riesgo particular en aquellas personas que utilizan *software de apoyo para la toma de decisiones* en análisis y diagnóstico (por ejemplo, médicos, gestores de riesgo o analistas financieros).

[107] Vid., *Directrices sobre decisiones individuales automatizadas*, cit., p. 28.

[108] Esta reforma del ET fue realizada a través del Real Decreto-ley 9/2021, de 11 de mayo, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales, que fue convalidado posteriormente por la Ley 12/2021, de 28 de septiembre (conocida como *Ley rider).* Se modifica el art. 64.4 ET para incluir un derecho de información colectiva, que se suma al derecho de información individual previsto en el RGPD, y que se aplica aun cuando se trate de *sistemas semiautomatizados* de toma de decisión. En virtud de este derecho, la representación legal de la plantilla tendrá que ser informada de *«los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que puedan incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles»*. Sobre este derecho señala GINÈS i FABRELLAS, A., *op.cit.,* p. 319 que, si bien se trata de una regulación pionera a nivel europeo que permite a la representación legal de la plantilla valorar la legalidad -e, incluso, potencial discriminatorio- de las variables utilizadas por el algoritmo, debería haberse acompañado con información estadística sobre el impacto del uso de dicha tecnología inteligente.

*lare*s en las personas, se trataría de que la decisión afectara de alguna manera a los derechos o la esfera jurídica de la persona, como por ejemplo sucede cuando te deniegan un préstamo o una subvención pública, te excluyen de un proceso de selección laboral o te despiden de un trabajo[109]. Es esencial, por tanto, que la decisión tenga el potencial de afectar significativamente a las circunstancias, al comportamiento o a las elecciones de las personas afectadas, impacte de forma prolongada o permanente en el interesado, o, en los casos más extremos, provoque la exclusión o discriminación de la persona[110].

Algún autor pone en duda que esta clase de decisiones (que a veces ni siquiera se refieren a personas concretas, sino a grupos o a zonas) produzcan los «efectos jurídicos en el interesado» a que se refiere el precepto[111]. Sin embargo, incluso en la conocida sentencia *Syri* del Tribunal de Distrito de La Haya, de 5 de febrero de 2020, se reconoce que los informes de riesgo elaborados por el sistema de IA acerca de la probabilidad de que determinados ciudadanos defraudaran al fisco, les afectaban «significativamente» de un modo similar a una decisión con efectos jurídicos.

A *sensu contrario*, las decisiones automatizadas basadas en datos personales que no tengan efectos o consecuencias jurídicas para los sujetos, no quedarían bajo el ámbito de aplicación del precepto analizado. Como ejemplo de este tipo de decisiones Herrera de las Heras pone el de un restaurante que utiliza un sistema de reconocimiento facial y detección de emociones para determinar el grado de satisfacción del cliente[112]. La decisión del sistema acerca del grado de satisfacción concreto de cada comensal no tendría efectos jurídicos para ellos en la medida en que no supondría que se les denegara el

---

[109] Pueden verse otros ejemplos de este tipo de decisiones en las *Directrices sobre decisiones individuales automatizadas*, cit., pp. 23-24. También en el Considerando 71 RGPD se pone como ejemplo la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna. Sin embargo, a pesar de dichos ejemplos, no resulta siempre evidente cuándo una decisión afecta significativamente a un sujeto. Como afirman MENDOZA, I./ BYGRAVE, L.A. «The Right Not to Be Subject to Automated Decisions Based on Profiling», *University of Oslo Faculty of Law Research Paper* n.20, 2017, p. 12, aunque puede deducirse que en aquellos casos en los que la decisión tenga impacto en la situación financiera o laboral del interesado estaremos ante un efecto significativo, el efecto concreto dependerá de las características de cada individuo, por lo que se deberá hacer un estudio *ad casum* sobre la significación o relevancia del efecto. Para EGUÍLUZ CASTAÑEIRA, J.U., *op.cit.,* pp. 342 y ss., un ejemplo de una decisión con «efectos jurídicos» sería una decisión judicial, o una decisión con respecto a un beneficio social otorgado por la ley, como el pago de la pensión. Por otro lado, un ejemplo de una decisión con efectos «significativamente similares» sería un banco que niega un crédito automáticamente. Y en base a lo que sostienen las Autoridades de Protección de Datos, la diferenciación de precios en línea podría «afectar significativamente de manera similar» a alguien, si conduce a «precios prohibitivamente altos que efectivamente excluyen a alguien de ciertos bienes o servicios».

[110] SÁEZ LARA, C., *op.cit.*, p. 53.

[111] Así, HUERGO LORA, A., «Una aproximación…», cit., p. 61.

[112] HERRERA DE LAS HERAS, R., *op.cit.,* p. 655.

acceso en próximas visitas, o que se les cobrara más o menos en función de su grado de satisfacción, sino que simplemente iría dirigida a evaluar su grado de satisfacción. En su opinión, parece que en estos casos el cliente no podría negarse al escrutinio hecho por el sistema inteligente. No obstante, a nuestro modo de ver, si bien no tendría los derechos y garantías del art. 22.3 ni el derecho de información reforzado de los arts. 13.2.f), 14.2.g) y 15.1.h), sí tendría derecho a ser informado del resto de extremos que contemplan tales normas y gozaría del resto de derechos que otorga la normativa cuando se procede al tratamiento de datos personales (acceso, rectificación, oposición…).

Por el contrario, un ejemplo de decisiones automatizadas con efectos jurídicos en la persona podría ser el rechazo automático de una solicitud de préstamo bancario basado únicamente en el perfil crediticio del solicitante generado por un algoritmo. Este tipo de decisión sí que tendría un efecto o consecuencia jurídica directa sobre la persona porque afectaría a su capacidad para obtener un crédito y, potencialmente, a su situación financiera general.

Presentes los dos presupuestos anteriores serían aplicables las garantías del art. 22.3 RGPD así como la regulación específica en relación al derecho de información y acceso del interesado previsto en los arts. 13.2 f), 14.2 g) y 15.1 h) [113], en cuya virtud, la persona afectada por la decisión tendrá *derecho a saber que está siendo objeto de una decisión automatizada*, es decir, derecho a conocer la existencia en sí del tratamiento automatizado de datos que le afectará, así como el derecho a que se le proporcione *información significativa sobre la lógica aplicada por el sistema y sobre la importancia y las consecuencias previstas de dicho tratamiento*. En consecuencia, los responsables estarían obligados a informar de ello (así como de su base legitimadora dada la prohibición general del art.22.1) y, al menos en tales casos, de los extremos anteriores. Por el contrario, no se aplicarían tales derechos y garantías si las decisiones no cumplieran los requisitos del art. 22, bien porque hubiera existido intervención humana significativa en su adopción (casos de decisiones semiautomatizadas), bien porque no desplegaran efectos jurídicos o similares sobre las personas.

Si bien esta conclusión era la que se venía manteniendo hasta el momento, la reciente sentencia del TJUE, de 7 de diciembre de 2023, ha sentado una interpretación distinta por lo que se refiere a los presupuestos del art. 22 RGPD, y en concreto, a su aplicación a las decisiones semiautomatizadas. Veámosla.

---

[113] Como explican las *Directrices sobre decisiones individuales automatizadas*, cit., del GT29, p. 28, «ofrecer esta información también ayudará a los responsables del tratamiento a garantizar que cumplen algunas de las garantías obligatorias descritas en el artículo 22, apartado 3, y el considerando 71».

c) La STJUE de 7 de diciembre de 2023, asunto C-634/21 (OQ y Land Hessen, con intervención de SCHUFA Holding AG)

En esta resolución el TJUE realiza una interpretación extensiva del art. 22 RGPD a fin de aplicar las garantías que contiene también a las decisiones parcial o semiautomatizadas[114]. El caso versaba sobre un litigio entre SCHUFA, una empresa alemana que proporciona información de solvencia a terceros, y OQ, un particular a quien un banco le denegó un préstamo en base a la información negativa que SCHUFA proporcionó sobre él. OQ ejerció su derecho de acceso a los datos registrados y su eliminación, pero SCHUFA se negó a revelar todos los datos utilizados en su evaluación. Solo cedió información genérica pero no los datos que tenía del afectado ni la ponderación de los mismos que había realizado en forma de valores de probabilidad. Justificó su negativa en que no había sido ella quien había denegado el crédito sino sus socios contractuales (es decir, el banco) e invocó también secretos comerciales[115].

El Tribunal de lo Contencioso-Administrativo de Wiesbaden planteó dos cuestiones prejudiciales al Tribunal de Justicia sobre la interpretación del art. 6 y 22.1 RGPD, con el fin de determinar, por un lado, si la actividad realizada por SCHUFA (esto es, la generación automatizada de un valor de probabilidad transmitido a terceros que toman decisiones basadas en él) constituía una «decisión automatizada» en el sentido del art. 22 RGPD teniendo en cuenta que no era SCHUFA quien tomaba la decisión final; y por otro lado, si la negativa de SCHUFA a proporcionar información adicional al afectado suponía una vulneración de los derechos de protección de datos de OQ por estar obligada a ello, teniendo en cuenta que los terceros que tomaron la decisión final -en base a dicho valor de probabilidad- no disponían de tal información.

---

[114] Sobre esta resolución puede verse, COTINO HUESO, L., «La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial» Diario LA LEY, Nº 80, Sección Ciberderecho, 17 de Enero de 2024. Recuperado de La Ley Digital, LA LEY 1301/2024. Para este autor «es reconfortante observar un progresivo reconocimiento normativo de la importancia de las garantías en decisiones parcial o semi-automatizadas», teniendo en cuenta que tanto la Carta de Derechos Digitales en España, como algunas legislaciones de protección de datos fuera de la UE también las recogen expresamente, como por ejemplo, la de Ecuador. También en Canadá o EEUU la definición de sistema de decisiones automatizado (*automated decision system*) incluye tanto las decisiones totalmente automatizadas como las de apoyo a la decisión. Vid., por ejemplo, el proyecto de Ley de Responsabilidad Algorítmica de 2022 de EEUU (sección 2; Definiciones).

[115] Las entidades de información y análisis de riesgo (como SCHUFA o en el caso del mercado español, Equifax o Experian) analizan el riesgo de un prestatario, combinando factores objetivos y subjetivos a partir de la información de que disponen sobre el prestatario, y/o de la información que sus socios comerciales (por ejemplo, los bancos) les proporcionan con el fin de que calculen esa puntuación o valor de probabilidad determinado. No obstante, son dichos socios comerciales los que toman la decisión final sobre la ejecución (o modo de ejecución) de un contrato con el prestatario, es decir, los que deciden la concesión o denegación del crédito, si bien con base en dicho valor.

En su sentencia, el Tribunal se detiene en el análisis de los presupuestos de aplicación del art. 22, que concreta en tres: en primer lugar, la existencia de una «decisión»; en segundo lugar, la necesidad de que ésta se base «únicamente» en un tratamiento automatizado de datos personales (incluida la elaboración de perfiles); y en tercer lugar, que la decisión despliegue «efectos jurídicos» en el interesado o le «afecte significativamente de modo similar».

Respecto al concepto de «decisión» al que se refiere el art. 22.1 RGPD, sostiene una interpretación extensiva en la línea mantenida por el Abogado General en el punto 38 de sus conclusiones. Entiende que puede incluir diversos actos con potencial para afectar al interesado de múltiples maneras y que engloba el resultado del cálculo de la solvencia de una persona en forma de un valor de probabilidad relativo a su capacidad para hacer frente a sus compromisos de pago en el futuro (párrafo 46).

Por lo que respecta al requisito de que la decisión deba estar «basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles», mantiene, como señalaba también el Abogado General en el punto 33 de sus conclusiones, que no se discute que una actividad como la que realiza SCHUFA responde a la definición de «elaboración de perfiles» del art.4.4 RGPD y, por tanto, que en el presente asunto se cumple este requisito, habida cuenta de que el tenor de la primera cuestión prejudicial se refiere expresamente a la generación automatizada de un valor de probabilidad a partir de datos personales relativos a una persona y acerca de su capacidad para satisfacer un préstamo en el futuro (párrafo 47).

Por último, respecto al requisito de que la decisión deba producir «efectos jurídicos» en el interesado o «[afectarlo] significativamente de modo similar», sostiene que del propio tenor de la primera cuestión prejudicial se desprende que la acción del tercero al que se transmite el valor de probabilidad (el banco) está basada «de un modo determinante» en dicho valor. Así pues, según las apreciaciones de hecho del órgano jurisdiccional remitente, en el caso de que un consumidor solicite un préstamo a un banco, un valor de probabilidad insuficiente dará lugar a que el banco deniegue la concesión del préstamo solicitado en la práctica totalidad de los casos (párrafo 48). En estas circunstancias, considera que la decisión sobre el valor de probabilidad, cuanto menos, afecta al interesado significativamente (párrafo 49) [116].

---

[116] Sobre las dificultades prácticas que pueden derivarse de la consideración de un valor como "determinante" o "no determinante", pueden verse las consideraciones de NEGRO, A., Y SANCHÉZ, A., «El veredicto del TJUE sobre el *credit scoring*», accesible en https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/veredicto-tjue-sobre-credit-scoring. Observan, entre otras cosas, que aunque el TJUE nada indica al respecto, cabe entender que la calificación de un valor como «determinante» no depende de que sea el único valor o puntuación que se tenga en cuenta para

Por tanto, el TJUE entiende que la generación automatizada del valor de probabilidad proporcionado por SCHUFA y utilizado por el banco para tomar decisiones sobre créditos, debe ser calificada como una decisión individual automatizada (basada en una elaboración de un perfil) que afecta significativamente al interesado. Ello implica la necesidad de cumplir con las disposiciones del RGPD sobre protección de datos y determina la obligación de SCHUFA de divulgar información adicional al interesado[117].

Esta interpretación se ve corroborada por el contexto en el que se inscribe el art. 22.1 RGPD, así como por los objetivos y la finalidad que persigue dicho Reglamento. Según entiende el TJUE, de no mantener esta interpretación existiría una laguna jurídica que mermaría la aplicación, finalidad y protección que se pretende con el art. 22 RGPD. Una interpretación restrictiva del precepto que excluyese decisiones como la controvertida (esto es, la generación automatizada de un valor de probabilidad) por considerarlas un mero *acto preparatorio* ejecutado por una agencia comercial, calificando únicamente como «decisión» del art. 22 el acto final adoptado por la tercera entidad, generaría un riesgo de elusión del precepto ya que los interesados no podrían ejercitar sus derechos frente a la empresa que calculó su probabilidad de impago, ni siquiera cuando (como ocurrió en este caso) el tercero que hubiera adoptado la decisión final no tuviera toda la información relevante acerca del tratamiento de datos realizado sobre interesado, al haber recibido, únicamente, el valor de probabilidad solicitado[118].

En definitiva, como explica Cotino Hueso, para el TJUE las garantías del art. 22 RGPD también se pueden aplicar a los actos, hechos o datos sobre los que se sustentan las decisiones automatizadas, como los perfiles o ponderaciones automatizados que han estado en la base de la decisión finalmente adoptada[119]. Ello a pesar de que *formalmente* la decisión final se adopte por

---

la toma de las decisiones pertinentes, *sino de que del valor trasladado haya dependido la decisión final* (la cursiva es nuestra).

[117] Esta conclusión la recoge el párrafo 73 en los términos siguientes: el art.22.1 del RGPD «debe interpretarse en el sentido de que la generación automatizada, por una agencia de información comercial, de un valor de probabilidad a partir de datos personales relativos a una persona y acerca de la capacidad de esta para hacer frente a compromisos de pago en el futuro constituye una "decisión individual automatizada", en el sentido de la mencionada disposición, cuando de ese valor de probabilidad dependa de manera determinante que un tercero, al que se comunica dicho valor, establezca, ejecute o ponga fin a una relación contractual con esa persona».

[118] Vid., párrafos 61 y 63.

[119] Ello supone, por un lado, que el responsable del tratamiento está sujeto a obligaciones adicionales de información en virtud de los artículos 13, apartado 2, letra f), y 14, apartado 2, letra g), de dicho Reglamento; y por otro lado, que el interesado tiene derecho, en virtud del artículo 15, apartado 1, letra h), del mencionado Reglamento, a obtener del responsable del tratamiento, en particular, «información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado» (párrafo 56). Estos requisitos más estrictos para la

una entidad distinta a la que hace el perfilado o ponderación previa, o de que haya aparentemente una intervención humana. Por ende, esta sentencia expande el alcance del art. 22 RGPD más allá del responsable formal de la decisión, para abarcar a aquellos que procesen datos y que efectúan *materialmente* el tratamiento automatizado para el responsable bajo otras posiciones jurídicas, como la de encargado por cuenta de quien finalmente toma la decisión[120].

Concluimos, junto al autor citado, que el criterio garantista que adopta el TJUE en esta resolución respecto de las decisiones automatizadas sienta un precedente que podrá proyectarse en muchos otros supuestos en el futuro. Precedente que, además, a nuestro juicio era totalmente necesario para subsanar el insuficiente ámbito de aplicación del art.22, teniendo en cuenta que, al menos en la práctica actual, es mucho más frecuente tomar decisiones con apoyo en los resultados de un sistema de IA, que delegar plenamente en ellos las decisiones finales[121].

---

licitud de una toma de decisiones automatizada, así como las obligaciones adicionales de información del responsable del tratamiento y los derechos de acceso adicionales del interesado correspondientes se explican por la finalidad que persigue el artículo 22 del RGPD, que consiste en proteger a las personas contra los riesgos específicos para sus derechos y libertades que supone el tratamiento automatizado de datos personales, incluida la elaboración de perfiles (párrafo 57).

[120] COTINO HUESO, L., «La primera sentencia del Tribunal de Justicia de la Unión Europea…», cit., se refiere a los sujetos intervinientes como «responsable formal de la decisión» (en este caso el Banco) y «tercero encargado del tratamiento de datos» (en este caso, Schufa) y considera que, en base a este pronunciamiento, las «garantías del artículo 22 RGPD se proyectan y van más allá del responsable -al menos formal- que toma la decisión respecto del afectado, generando obligaciones a un tercero encargado que ha tratado datos del afectado». No obstante, para otros autores, las agencias como Schufa que prestan servicios de *credit scoring* actuarán directamente en calidad de «responsables del tratamiento» cuando concurran los tres requisitos analizados por el TJUE y precisamente por ello deberán cumplir las obligaciones que el RGPD impone a los mismos: por un lado, asegurar la aplicación de alguna de las excepciones previstas en el artículo 22.2 y contar con una base legitimadora el tratamiento de los datos de los prestatarios que deberán probar. Asimismo, deberán garantizar el cumplimiento de la normativa aplicable (por ejemplo, aplicando las medidas técnicas y organizativas apropiadas para la seguridad de la información y la reducción del riesgo de error al máximo o garantizar los derechos de los interesados). Así lo mantienen NEGRO, A., Y SANCHÉZ, A., *op.cit.loc.cit.* Sea como fuere, en ambos casos los autores coinciden en que los sujetos que *materialmente* efectúan el tratamiento automatizado, perfilado o ponderación de datos quedarían sujetos a ciertas obligaciones.

[121] No obstante, como augura la doctrina, esta situación parece que se revertirá en un futuro no muy lejano. En este sentido, afirma PALMA ORTIGOSA, A., *op.cit.,* p. 6/25, que «el paso del tiempo hará que las decisiones automatizadas que ahora suelen utilizarse como un parámetro más de apoyo a la hora de tomar una decisión determinada, se conviertan en la regla a seguir, pudiendo llegar un momento de estandarización donde se pase de la desconfianza de la máquina, al recelo de la decisión que pueda adoptar un humano contraria a la que ha indicado el algoritmo, de manera que ya no se dude de la máquina sino de la persona».

4.2.2. Contenido del derecho de información de los arts. 13.2 g), 14.2 g) y 15.1 h)

El RGPD no concreta qué implica el deber de proporcionar «*información significativa sobre la lógica aplicada*» y si dicha expresión conlleva o no el derecho a conocer el código fuente del algoritmo y su funcionamiento. Dicha omisión ha hecho que surjan dudas y vacilaciones acerca de esta cuestión[122]. Es opinión ampliamente compartida que esa disposición del RGPD no impone una obligación de informar y dar acceso de forma exhaustiva al código fuente del algoritmo utilizado para adoptar la decisión, lo que, por otra parte, tampoco sería de utilidad en muchos casos para el interesado si no se comparten también los conjuntos de datos utilizados, por cuanto que en los sistemas adaptativos el código fuente va cambiando[123]. Según las *Directrices sobre deci-*

---

[122] Sobre este asunto planteaba RUBI PUIG dos posibilidades: que la información a suministrar tuviera que consistir en una explicación general del modelo (*model-based explanation*) o en explicar cómo se adoptó una decisión basada en los datos particulares del afectado (*subject-based explanation*). Vid., «Elaboración de perfiles y personalización de ofertas y precios en la contratación con consumidores», *Revista De Educación Y Derecho*, 24, 2021, https://doi.org/10.1344/REYD2021.24.36304, p. 17/24. Sobre esta doble posibilidad, aclara MEDINA GUERRERO, M., *op.cit.,* p. 161, que el hecho de que la información deba facilitarse al recabarse los datos o en un plazo breve de tiempo después (ex arts. 13.2 y 14.3a) hace que en esta fase (previa a la toma de decisión) la obligación de informar no pueda traducirse en la exigencia de asegurar la explicabilidad, ni la fundamentación de una decisión en particular. Lo que debe ponerse en conocimiento del interesado es *la descripción de los tipos de datos y los factores que toma en consideración el sistema, las categorías del árbol de decisión y la ponderación de tales factores a nivel global* (la cursiva es nuestra). Con esa información el afectado debe poder prever qué papel juegan (o pueden jugar) sus diferentes datos objeto de tratamiento en la formación de la decisión a tomar, así como estar en condiciones de acomodar su conducta a los criterios que resultan determinantes para adoptar la decisión. Para NAVAS NAVARRO, S., «La perspectiva de género…», *cit.,* se puede interpretar que la lógica de la que se debe informar en estos casos es la lógica computacional empleada y se trataría de que se pudiera conocer, en la medida de lo posible, el proceso de toma de decisiones para ver si existe algún sesgo o error en el procesamiento de los datos.

[123] Según concluye el GT29 en sus *Directrices sobre decisiones individuales automatizadas*, p. 28, dicho deber de proporcionar información no consiste en revelar todo el algoritmo y menos aún su código fuente, o el secreto empresarial que lo protege. A la misma conclusión llega, por ejemplo, el MINISTERIO DE TRABAJO Y ECONOMÍA SOCIAL*, Información algorítmica en el ámbito laboral. cit.,* pp. 12-15, según el cual, «la obligación de información derivada de los artículos 13.2.f) y 14.2.g) 15.1.h) RGPD y 64.4.d) ET no puede interpretarse como la obligación empresarial de facilitar el código fuente del algoritmo». Comparten esta postura, RIVAS VALLEJO, P., «Herramientas desde el derecho antidiscriminatorio…», *cit.,* p. 552, NÚÑEZ SEOANE, J., *op.cit.*, p. 307, o LAZCOZ MORATINOS, G., *op.cit.,* entre otros. Para la mayoría de la doctrina, este derecho de información no implica el derecho a acceder al algoritmo, no solo porque puede estar protegido por secreto empresarial, sino porque implicaría acceder a centenares de páginas de códigos indescifrables. Por el contrario, como precisa GINÈS i FABRELLAS, A., *op.cit.,* p. 319, implica el derecho a obtener información sobre las variables o métricas utilizadas por el algoritmo para la elaboración de perfiles o toma de decisiones automatizada, su ponderación en la ecuación final y las consecuencias que pueden derivarse para la persona. Como explica PALMA ORTIGOSA, A., *op.cit.,* p.14/25, el responsable del tratamiento debe explicar, en la medida de lo posible, el funcionamiento del programa informático, aportando información útil para el interesa-

siones individuales automatizadas del GT29, se trataría de ofrecer información conforme al *principio de transparencia* en términos sencillos y comprensibles, pero suficientemente exhaustiva para que el interesado pudiera entender *los motivos de la decisión,* sin necesidad de incluir una explicación compleja sobre los algoritmos utilizados, lo que podría ser opaco, confuso, e incluso conducir a la fatiga informativa[124]. A nuestro modo de ver, el *quid* de la cuestión sería que la información proporcionada permitiera al interesado hacer efectivos los derechos de los que es titular. Por tanto, creemos junto con un sector doctrinal, que dicha información no sería suficiente ni significativa si impidiera al interesado ejercitar los derechos y garantías que el propio art. 22 RGPD le otorga, esto es, oposición, revisión, alegación e impugnación de la decisión[125].

Sobre el contenido concreto de este derecho es interesante la opinión del Abogado General Priit Pikamäe en el asunto C-634/21 y en los asuntos acumulados C-26/22 y C-64/22 SCHUFA Holding y otros, al que hicimos referencia en el apartado anterior. Considera que *no es suficiente con dar una explicación genérica sobre cómo funciona el proceso de toma de decisión, sin indicar los datos específicos que se tienen de la persona y que han sido tenidos en cuenta para elaborar el informe negativo de solvencia.* A su juicio, «la obligación de proporcionar "información significativa sobre la lógica aplicada" debe entenderse en el sentido de que incluye *explicaciones suficientemente* detalladas sobre el método utilizado para el cálculo del *score* y sobre las *razones que han conducido a un resultado determinado.* En general, el responsable del tratamiento debería proporcionar al interesado información general, en particular sobre los *factores que han sido tenidos en cuenta en el proceso de toma de decisiones y sobre su importancia relativa desde el punto de vista agregado, que también le resulte útil para impugnar cualquier "decisión" en el sentido de la disposición del RGPD* que consagra el "derecho" a no ser objeto de

---

do - como por ejemplo los factores que más han influido en la elaboración del perfil - que le permita comprender el proceso y pueda así ejercer el resto de los derechos que le reconoce el RGPD en su artículo 22.3. En la *Directiva (UE) 2019/2161 del Parlamento Europeo y del Consejo de 27 de noviembre de 2019 por la que se modifica la Directiva 93/13/CEE del Consejo y las Directivas 98/6/CE, 2005/29/CE y 2011/83/UE del Parlamento Europeo y del Consejo,* sobre la mejora de la aplicación y modernización de las normas de protección de los consumidores de la Unión, también se aclara que no debe exigirse a los comerciantes que revelen el funcionamiento detallado de sus mecanismos de clasificación (incluidos los algoritmos), sino sólo una descripción general de los principales parámetros que determinan la clasificación, si bien no es necesario que dicha descripción se presente de un modo personalizado para cada una de las consultas concretas efectuadas (vid., Considerandos 22 y 23).

[124] La AEPD señala que según la RAE la palabra «*significativa*» denota «*Que da a entender o conocer con precisión algo*» y esto habría que interpretarlo como información que, proporcionada al interesado, le hace consciente del tipo de tratamiento que se está llevando a cabo con sus datos y que le proporciona certeza y confianza sobre sus resultados, lo que está relacionado con el concepto de «explicabilidad» del tratamiento mediante IA.

[125] En este sentido, por ejemplo, NÚÑEZ SEOANE, J., *op.cit.*, p. 307 y NAVAS NAVARRO, S., «La perspectiva de género…», *cit.,* p. 15/23.

una decisión basada únicamente en un tratamiento automatizado, incluida la elaboración de perfiles»[126].

Sin perder de vista las premisas y consideraciones anteriores, habría que aclarar sobre qué extremos concretos se debería informar al interesado. Tal como indica la AEPD, aunque la respuesta dependerá del tipo de componente de IA utilizado, un ejemplo de información que podría tener relevancia para el interesado podría ser la siguiente: *«El detalle de los datos empleados para la toma de decisión, más allá de la categoría, y en particular información sobre los plazos de uso de los datos (su antigüedad); la importancia relativa que cada uno de ellos tiene en la toma de decisión; la calidad de los datos de entrenamiento y el tipo de patrones utilizados; los perfilados realizados y sus implicaciones; los valores de precisión o error según la métrica adecuada para medir la bondad de la inferencia; la existencia o no de supervisión humana cualificada; la referencia a auditorías, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como la certificación o certificaciones realizadas sobre el sistema de IA. En el caso de sistemas adaptativos o evolutivos, la última auditoría realizada; en el caso de que el sistema IA contenga información de terceros identificables, la prohibición de tratar esa información sin legitimación y de las consecuencias de realizarlo»*[127].

Por su parte, el Ministerio de Trabajo y Seguridad Social de España publicó en 2022 una *Guía práctica sobre la obligación de los empresarios de informar a los trabajadores sobre el uso de algoritmos*[128] que, aunque dirigida al ámbito de las relaciones laborales, creemos extrapolable a cualquier otro, al menos en lo relativo al contenido concreto de la información que deberá proporcionarse al interesado respecto del sistema algorítmico utilizado para la toma de decisiones que le afectan. En dicho documento se establece, con gran detalle, que deberá informarse de los siguientes extremos:

a) *Información sobre el uso de algoritmos o sistemas de IA* para tomar decisiones automatizadas, incluyendo la elaboración de perfiles, identificando la tecnología utilizada por parte del algoritmo y las decisiones respecto de las que se utiliza tal tecnología. Esto incluiría, por ejemplo, informar sobre si la tecnología empleada genera un algoritmo de «caja negra» o si se trata de un algoritmo de aprendizaje continuo, el concreto *software* o producto utilizado y si cuenta, en su caso, con algún tipo de certificación, la empresa suministradora, si la em-

---

[126] Comunicado de prensa del TJUE n. 49/2023, 16 de marzo de 2023, con las conclusiones del Abogado General disponible en: https://curia.europa.eu/jcms/jcms/Jo2_7052/es/?annee=2023. Las cursivas del texto son nuestras.

[127] P. 24/53.

[128] Dicha Guía, que fue elaborada por una comisión formada por expertos en algoritmos, concreta y sistematiza las obligaciones empresariales de información con el fin de garantizar, entre otros, el derecho fundamental a la no discriminación.

presa ha realizado alguna alteración sobre el producto, o el grado de intervención humana cualificada en las decisiones adoptadas, esto es, si las mismas se adoptarán mediante sistemas de IA o simplemente con apoyo en los resultados de dichos sistemas.

b) *Información significativa, clara y simple sobre la lógica y funcionamiento del algoritmo*, que incluiría tanto *las variables tomadas en consideración* (entendidas como la información o factores utilizados por el algoritmo para tomar la decisión o la elaboración del perfil[129], incluyendo si algunas de estas variables son datos personales), como *los parámetros utilizados* (entendidos como la ponderación relativa de cada variable en el modelo para la toma de la decisión, así como cualquier cambio de estos parámetros que modifique el comportamiento del algoritmo)[130]. También habría que informar sobre las *reglas e instrucciones utilizadas por el algoritmo,* entendidas como las reglas programación (ya sea las programadas de forma expresa o las derivadas del aprendizaje del algoritmo) que conducen a la toma de la decisión[131] y sobre los *datos de entrenamiento* y, en su caso, *validación* (ya que estos también influyen en su lógica o instrucciones del algoritmo), su calidad (adecuación, pertinencia, minimización atendiendo a la finalidad para los que fueron obtenidos, etc.) y el *tipo de patrones identificados en los datos de entrenamiento*[132]. En casos de elaboración de perfiles, las métricas de precisión o error en las tareas automatizadas (clasifica-

[129] En caso de elaboración de perfiles, también hay que informar de la tipología de perfiles que elabora el algoritmo y, específicamente a la persona trabajadora ex artículos 13.2.f) y 14.2.g) RGPD, habría que darle *información sobre la asignación concreta a uno de ellos.*

[130] Por ejemplo, en un sistema algorítmico utilizado en un proceso de selección, las variables podrían ser la formación de la persona o la experiencia profesional previa, y los parámetros serían el peso relativo que tiene cada variable en la decisión final de selección o contratación (v.gr. el modelo podría ordenar a las personas en atención a su experiencia profesional previa -valorada en un 40%-, formación -valorada en un 20%-etc., hasta alcanzar el 100%).

[131] En un sistema IA utilizado por una plataforma de reparto de comida a domicilio, las reglas e instrucciones podrían ser que la asignación de pedidos entre sus repartidores se realizara de la siguiente manera: (i) excluyendo a las personas que se encuentran a una distancia superior de 1 km del restaurante, (ii) puntuando en una escala de X a Y en un 50% a las personas que se encuentran a una distancia mínima de 1 km del restaurante y en otro 50% a las personas con una disponibilidad mínima de 10 horas en alta demanda en las últimas dos semanas y (iii) asignando el pedido a la persona que obtuviera una puntuación superior; así como las demás reglas de programación, tales como aplicar las anteriores instrucciones a las personas conectadas en ese momento, excluir a las personas que están en un pedido en curso, etc.

[132] Por ejemplo, en un algoritmo para un proceso de selección la empresa tendría que comunicar que ha utilizado (i) los datos de la plantilla de la empresa de los últimos 10 años; (ii) obtenidos cumpliendo con las exigencias de protección de datos, específicamente la obligación de haber informado a estas personas que sus datos serían reutilizados para el entrenamiento de un algoritmo para elaborar perfiles; y (iii) informar que, analizando los datos de entrenamiento, ha identificado el patrón estadístico que, por ejemplo, las personas con una formación en A cumplen un X% mejor

ción, puntuación, regresión, ordenación…) de las personas en los distintos perfiles. Y, por último, la referencia a las *auditorías* realizadas (en el caso de sistemas adaptativos o evolutivos la última auditoría realizada) *especialmente sobre las posibles desviaciones de los resultados de las inferencias,* o la *evaluación de impacto* realizada por parte de la empresa o una tercera empresa respecto del algoritmo o sistema de decisión automatizada utilizado.

c) Información sobre las *consecuencias que puede tener sobre la persona la decisión* adoptada (por ejemplo, obtener un puesto de empleo o determinar las concretas condiciones laborales) [133].

En los casos en que las decisiones automatizadas exigieran la previa realización de una evaluación de impacto (art. 35 RGPD), el contenido de la información suministrada podría ser aún más exhaustivo, ya que sería ésta la que determinaría las «medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado» [134].

Toda la información suministrada debería de cumplir, además, las exigencias formales y el nivel de transparencia del art. 12 RGPD, esto es, habría de facilitarse de forma clara, simple y comprensible a fin de que el interesado pudiera ejercer sus derechos.

Para garantizar la efectividad de dichos derechos, el RGPD confiere poderes de investigación y sanción a la autoridad de control (vid., arts. 53 y 83). Para un sector de la doctrina, esta correlación de derechos individuales y de

---

sus funciones para un puesto determinado, en comparación con aquellas personas que tienen una formación en B.

[133] Además, en el ámbito laboral, debe informarse a los representantes de los trabajadores sobre el impacto que las decisiones adoptadas mediante algoritmos o sistemas de decisión automatizada tienen en materia de igualdad y no discriminación entre mujeres y hombres. Así se deriva del artículo 64.3 ET que reconoce el derecho de la representación legal de la plantilla de acceder a información referente a la aplicación en la empresa del derecho a la igualdad y no discriminación entre mujeres y hombres. Así como del artículo 7 Real Decreto 901/2020, referente al diagnóstico en un contexto de elaboración de un Plan de Igualdad.

Vid., AEPD., *Adecuación…, cit.,* pp. 24 y ss., en relación con el alcance de la información sobre la «*importancia*» y «*consecuencias previstas*» apunta que la misma debe versar «sobre el tratamiento previsto o futuro, y sobre cómo puede afectar la decisión automatizada al interesado».

[134] Para NÚÑEZ SEOANE, J., *op.cit.*, pp. 312-313, la evaluación de impacto es un recurso alternativo para que el responsable del tratamiento que por legítimos motivos no quiera o no pueda revelar el funcionamiento del algoritmo, pueda justificar y demostrar ante el interesado con plenas garantías que las decisiones automatizadas que adopte no supondrán un riesgo para sus derechos y libertades. Para RIVAS VALLEJO, P., «Herramientas desde el derecho antidiscriminatorio…», *cit.,* p. 553, a diferencia de las auditorías del algoritmo, la evaluación de impacto permite actuar igualmente con carácter previo a la implantación de decisiones normativas o privadas que comporten el uso de algoritmos de decisión con el fin de evitar su impacto futuro y constituiría una barrera adicional a la auditoría de la IA y a la participación de la representación legal de los trabajadores ex art. 65.5 ET.

poderes o competencias de la AEPD es lo que realmente garantiza el cumplimiento de los deberes de *transparencia hacia el interesado,* y por ende, la efectividad del derecho a una explicación de las personas afectadas por las decisiones automatizadas[135].

Llegados a este punto, y en base a todos estos mimbres, se puede concluir que el RGPD se erige en uno de los principales instrumentos para hacer frente a la opacidad de los sistemas de toma de decisiones con impacto en las personas. Por un lado, porque trata de impedir el *primer nivel de opacidad* al que nos referíamos *supra* cuando establece en los arts.13.2 f), 14.2 g) y 15.1 h) el derecho a ser informado de la *existencia de decisiones automatizadas*[136]. Así, los interesados no deberían ignorar, en teoría, que el proceso decisorio que les afecta se va a realizar por medio de un tratamiento automatizado de sus datos personales. El problema reside en que, en la práctica, no llegan a ser conscientes de ello porque los responsables del tratamiento no les informan de este extremo al considerar que su decisión no es automatizada, aunque de hecho sí lo sea[137]. Por otro lado, porque pretende evitar el *segundo nivel de opacidad*, ya que en tales casos deberá proporcionarse al interesado información significativa con el contenido y alcance anteriormente comentados. Por ende, puede afirmarse que en cierta medida dicha normativa contribuye a que el interesado pueda desafiar la validez de un algoritmo y pueda acceder a la ponderación que se le otorga a sus diversos elementos y a cualquier conclusión errónea a la que llegue.

Sin perjuicio de lo afirmado, la efectividad del derecho de información reconocido en el RGPD y su correlativo deber de transparencia, ha sido puesta en duda por la doctrina en los casos en que se emplean sistemas automatizados de decisiones basados en técnicas de IA como el *machine o deep learning*. Ello por diversas razones: por un lado, porque cuando se trata de algoritmos de *aprendizaje automático* que van modificándose progresivamente y de forma autónoma, es difícil proporcionar la información sobre la lógica del algoritmo por el efecto de caja negra que los caracteriza; por otro lado, porque como se explicó en su momento, en muchos casos se alegan derechos de propie-

---

[135] En este sentido, SÁEZ LARA, C., *op.cit.*, pp. 52 y ss.

[136] LAZCOZ MORATINOS, G., *op.cit.,* p. 293.

[137] Advierte de este hándicap PALMA ORTIGOSA, A., *op.cit.,* p. 4/25, para quien, la indeterminación de los conceptos jurídicos recogidos en el art. 22 RGDP, conlleva que en muchas ocasiones los responsables desconocerán de buena voluntad si se trata de un tratamiento al que deban aplicarse las reglas del precepto. Además, teniendo en cuenta que este precepto establece mayores exigencias para ellos y más derechos en favor de los titulares de los datos, ante situaciones dudosas tenderán a huir de su aplicación por entender que el tratamiento que están llevando a cabo no se asimila al definido en el precepto. Esclarecer todos estos conceptos jurídicos ayudaría al conjunto de operadores jurídicos a determinar cuándo el tratamiento concreto que se está llevando a cabo se subsume en el art.22 y cuándo no.

dad intelectual o secreto empresarial para no ofrecer información algorítmica a los interesados, no habiendo una postura doctrinal unánime acerca de la prevalencia o no de dichos derechos sobre los derechos o intereses de los individuos afectados por el tratamiento de sus datos personales cuando ambos entran en conflicto[138]. Estas y otras razones llevan a un sector doctrinal a mantener que si bien el art. 22 RGPD es útil en el contexto de la protección de datos, no lo es tanto para el control del uso de la IA aplicada a datos masivos[139].

No obstante, a nuestro modo de ver, en la medida en que hay consenso en que sería suficiente con compartir una explicación funcional en lenguaje común para que cualquier persona interesada, sin los conocimientos técnicos suficientes, pudiera entender bajo qué parámetros ha podido ser afectado por un modelo matemático de decisión, sin necesidad de revelar el código fuente o el secreto empresarial que lo protege, se evitarían en gran medida perjuicios a los derechos de propiedad intelectual o secretos comerciales y la negativa a proporcionar información carecería de justificación. Más aún si se tiene en cuenta que los expertos mantienen que existen métodos en la ciencia de la computación que permiten supervisar el modelo para averiguar si es necesario hacer ajustes o correcciones que no comprometen el contenido del algoritmo cuando está sujeto a derechos de propiedad intelectual o incluso por razones de seguridad[140].

4.3. La transparencia y la explicabilidad en el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial

La transparencia ocupa un lugar destacado en el RIA y se refiere tanto al grado de implantación en un sistema (finalidad, diseño, función y datos)

---

138 Sobre esta polémica puede verse LÓPEZ-TARRUELLA MARTÍNEZ, A., *op.cit.*, pp. 197-199.

139 Así lo entiende REBOLLO DELGADO, L., *op.cit.,* pp. 108, para quien «dicho precepto aporta una solución tangencial a un problema sobrevenido, como es la aplicación de la IA a bases masivas de datos, por lo que su eficacia es limitada. Parece necesaria una regulación más específica, que concrete todas las posibilidades y lo haga desde la perspectiva de la protección del usuario, de la persona implicada. Pero aquí nos volvemos a encontrar con el problema recurrente de qué tipo de información ofrecemos, qué nivel técnico tiene, y en qué grado es entendible para el usuario o titular de los datos. Este problema de aplicabilidad no es menor, y plantea la vuelta a la dinámica de la Directiva 95/46/UE y que desterró el RGPD, es decir, que al responsable o encargado del tratamiento no le es suficiente con cumplir la norma, sino que debe conseguir la finalidad propuesta por ella. Esto supone un endoso al responsable o encargado del tratamiento de la teleología del mandato jurídico de la norma, cuya efectividad suele quedar difuminada».

140 RIVAS VALLEJO, P., «Herramientas desde el derecho antidiscriminatorio…», *cit.,* p. 555, citando estudios sobre el tema.

como al modo en que se utiliza y a las repercusiones que genera[141]. Además, aunque el RIA aboga principalmente por la *transparencia interna* (sobre todo hacia los responsables del despliegue), también establece algunas obligaciones a favor de la *transparencia externa* y ha acabado incorporando, como novedad destacable respecto a su versión inicial, el derecho a una explicación de las personas afectadas por las decisiones algorítmicas.

4.3.1. Obligaciones de transparencia y comunicación de información

A) Sistemas de alto riesgo

Cuando los sistemas no son lo suficientemente transparentes y explicables, ni están bien documentados, pueden impedir el ejercicio de derechos procesales fundamentales, como el derecho a la tutela judicial efectiva y a un juez imparcial, el derecho de defensa y la presunción de inocencia[142]. Para evitar la conculcación de tales derechos, el RIA establece, en su art. 13.1, que los sistemas de alto riesgo «se diseñarán y desarrollarán de un modo que garantice que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente su información de salida. Se garantizará un tipo y un nivel de transparencia adecuados para que el proveedor y el responsable del despliegue cumplan las obligaciones oportunas previstas en la sección 3». De esta disposición pueden extraerse varias ideas:

En primer lugar, en cuanto al *grado de transparencia* requerido para subsanar la opacidad de algunos sistemas, se exige una transparencia *suficiente* y no total[143], con el fin de evitar colisiones con los derechos de propiedad intelectual. Con esta declaración, la Comisión se posiciona claramente a favor del

---

141 FERNÁNDEZ ROZAS, J.C., *op.cit.*, p. 25.

142 Así se advierte en el Considerando 59 RIA.

143 Para HUERGO LORA, A. «El proyecto de Reglamento sobre la Inteligencia Artificial», 2021. Disponible en https://almacendederecho.org/el-proyecto-de-reglamento-sobre-la-inteligencia-artificial, la regulación de la transparencia es una de las cuestiones más delicadas, y el proyecto busca un equilibrio en virtud del cual el proveedor debe mostrar cómo funciona la aplicación, incluida su «lógica general», así como los presupuestos de partida o una descripción de los datos que se hayan utilizado para su creación, pero no se le exige transparencia total sobre el *software* utilizado. Considera el autor, que «estos requisitos son, en cierto modo, "objetivos" máximos a los que se debe tender, pero que pueden conseguirse de muchas maneras y también con niveles de intensidad diferentes». En este sentido se pronuncia la *Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO* (2021), al considerar que el grado de transparencia y explicabilidad debería ser siempre adecuado al contexto y al efecto, ya que puede ser necesario encontrar un equilibrio entre la transparencia y la explicabilidad y otros principios como la privacidad, la seguridad y la protección (ap. 38).

equilibrio entre la necesidad de transparencia y los intereses empresariales, en la misma línea que ya venía manteniendo el Consejo de Europa. La versión inicial del RIA indicaba expresamente en su Exposición de Motivos que *ni siquiera las obligaciones que exigían una mayor transparencia debían afectar de manera desproporcionada a los derechos de propiedad intelectual* puesto que únicamente debían aplicarse a la información mínima necesaria para que las personas pudieran ejercer su derecho a una compensación efectiva y solo exigían la transparencia necesaria hacia las autoridades de supervisión y las encargadas de la aplicación de la ley. También se establecía que la información habría de divulgarse siempre con arreglo a la legislación pertinente en la materia, y que, tanto las autoridades públicas como los organismos notificados tendrían que cumplir obligaciones de confidencialidad vinculantes en los casos en que debieran tener acceso a información confidencial o al código fuente para examinar el cumplimiento de las obligaciones sustanciales[144]. En la versión definitiva del RIA también se establece reiteradamente que toda la información o documentación a la que tengan acceso las autoridades, organismos notificados y cualquier otra persona física o jurídica que participe en su aplicación, deberá ser tratada de conformidad con las obligaciones de confidencialidad establecidas en el artículo 78.

En segundo lugar, el *tipo y nivel de transparencia* exigible ha de ser compatible con el cumplimiento de las obligaciones legales de proveedores y responsables del despliegue.

En tercer lugar, el RIA *asocia la transparencia con la información que se debe suministrar al responsable del despliegue del sistema.* Es decir, en principio la obligación de transparencia y comunicación de información va dirigida a informar a los agentes que implementan la tecnología y no a las personas potencialmente afectadas por la decisión automatizada. Por tanto, se trata de una transparencia interna y no hacia los usuarios finales.

El Considerando 72 RIA da amplia cuenta de que el objetivo de la transparencia exigible a los sistemas de alto riesgo es que los responsables del despliegue puedan «*comprender* la manera en que el sistema de IA funciona, evaluar su funcionalidad y comprender sus fortalezas y limitaciones», así como «*utilizar el sistema y tomar decisiones con conocimiento de causa*». No obstante, como se explicará *infra*, esto repercutirá indirecta y favorablemente en la transparencia hacia los usuarios finales -transparencia externa- en la medida en que contribuirá a dar cumplimiento a las exigencias del art. 86 que obliga a los responsables a dar explicaciones a los afectados por la decisión del sistema.

---

[144] Apartado 3.5 de la Exposición Motivos que contenía la Propuesta de RIA de 21 de abril de 2021 -COM(2021) 206 final-.

Con estos fines, los sistemas de IA de alto riesgo deben ir acompañados de la documentación y las instrucciones de uso oportunas (en un formato digital o de otro tipo adecuado) e incluir información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los responsables del despliegue, en particular sobre los posibles riesgos para los derechos fundamentales y de discriminación, cuando corresponda. Concretamente, en las instrucciones de uso el proveedor de la tecnología quedará obligado a informar al responsable del despliegue de forma adecuada y suficiente sobre una serie de extremos *mínimos* que se recogen en el art. 13.3, entre los que se incluyen, por ejemplo, las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo. Además, para facilitar que los responsables del despliegue comprendan dichas instrucciones, deberán contener ejemplos ilustrativos según proceda[145].

Por otro lado, otro requisito exigible por el RIA que también contribuye a la transparencia externa, es que los sistemas de alto riesgo deban *registrarse en una base datos de la UE* que la Comisión gestionará con el propósito de redoblar la transparencia y vigilancia públicas y de fortalecer la supervisión *ex post* por parte de las autoridades competentes (art. 49 en relación con el art. 71)[146]. Como dispone el art. 71.4, la información de esta base de datos registrada de conformidad con lo dispuesto en el artículo 49 será accesible y estará a disposición del público de manera sencilla.

Por último, aunque no serán objeto de análisis en este trabajo, debe al menos mencionarse que los proveedores e implantadores de *modelos de IA de uso general*, quedan sometidos a obligaciones de transparencia más exigentes en virtud del RIA.

B) Determinados sistemas de IA con independencia de su riesgo

El art. 50 impone a proveedores y responsables del despliegue obligaciones de transparencia específicas respecto a ciertos sistemas de IA, con independencia de si son o no de alto riesgo[147]. Se trata de aquellos destinados a interactuar directamente con personas físicas, sistemas que generen contenido sintético de audio, imagen, vídeo o texto, sistemas de reconocimiento de

---

145 Considerando 72.

146 Que el objetivo de esta obligación es aumentar la transparencia hacia el exterior o de cara al público lo expresa claramente el Considerando 131: «Con el objetivo de facilitar la labor de la Comisión y de los Estados miembros en el ámbito de la IA, así como de incrementar la transparencia de cara al público …».

147 En los Considerandos 132, 133 y 134 se justifica la imposición de tales obligaciones a estos sistemas.

emociones o de categorización biométrica y sistemas que generen o manipulen imágenes o contenidos de audio o vídeo que constituyan una ultrafalsificación, o generen o manipulen texto que se publique para informar sobre asuntos de interés público. En todos estos casos, se exceptúan de esta obligación los sistemas de IA autorizados legalmente para detectar, prevenir, investigar o enjuiciar delitos.

El objetivo que se persigue en estos casos con la transparencia e información es evitar la confusión o engaño de quienes interactúan o se ven expuestos a dichos sistemas[148]. Para ello, en el primer caso, se deberá comunicar a las personas que están interactuando con un sistema de IA y no con un humano (excepto cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización). Al aplicar dicha obligación, deben tenerse en cuenta las características de las personas físicas pertenecientes a colectivos vulnerables debido a su edad o discapacidad, en la medida en que el sistema de IA esté destinado a interactuar también con dichos colectivos.

En el segundo caso, se velará por que la información de salida del sistema esté marcada en un formato legible por máquina y que sea posible detectar que ha sido generada o manipulada de manera artificial y no por un ser humano. Se exige a los proveedores que integren soluciones técnicas y métodos para lograr este cometido que deberán ser lo suficientemente fiables, interoperables, eficaces y sólidos, en la medida en que sea técnicamente viable, teniendo en cuenta las técnicas disponibles o una combinación de dichas técnicas, como marcas de agua, identificación de metadatos, métodos criptográficos para demostrar la procedencia y la autenticidad del contenido, métodos de registro, impresiones dactilares u otras técnicas, según proceda. Además, a la hora de aplicar esta obligación, deberán tener en cuenta las particularidades y limitaciones de los diversos tipos de contenido, los costes de aplicación y el estado actual de la técnica generalmente reconocido, según se refleje en las normas técnicas pertinentes[149].

---

[148] Señala FERNÁNDEZ ROZAS, J.C., *op.cit.*, p. 16, que con los requisitos de transparencia que se imponen a los sistemas que interactúan con seres humanos se pretende dar lugar a un cambio significativo en las prácticas de divulgación a escala global en sitios web y aplicaciones, y garantizará que los usuarios estén informados sobre el funcionamiento y el propósito de los sistemas de IA con los que interactúan.

[149] Indica el Considerando 133 que «Dichas técnicas y métodos pueden implantarse a nivel de sistema de IA o a nivel de modelo de IA, incluidos modelos de IA de uso general que generan contenidos, facilitando así el cumplimiento de esta obligación por parte del proveedor posterior del sistema de IA. Para garantizar la proporcionalidad, conviene prever que esta obligación de marcado no se aplique a los sistemas de IA que desempeñen una función de apoyo a la edición estándar o no alteren sustancialmente los datos de entrada facilitados por el responsable del despliegue o su semántica».

En el tercer caso, se notificará a las personas físicas que están expuestas a sistemas de IA que tratan sus datos biométricos y que pueden determinar o inferir sus emociones o intenciones, así como incluirlas en categorías específicas en función del sexo, edad, color del pelo o de ojos, tatuajes, rasgos personales, origen étnico o preferencias e intereses personales. Dichos datos personales se tratarán de conformidad con el Reglamento y la Directiva de protección de datos personales. Además, si las personas cuyos datos se tratan son personas con discapacidad, toda la información y notificaciones deberán facilitarse en formatos accesibles.

Por lo que se refiere al último tipo de sistemas mencionado, deberá hacerse público que los contenidos, imágenes o texto han sido generados o manipulados de manera artificial etiquetando los resultados de salida generados por la IA e indicando su origen artificial. No obstante, cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, sólo se deberá hacer pública la existencia de dicho contenido generado o manipulado artificialmente de una manera adecuada que no dificulte la exhibición o el disfrute de la obra, su explotación y uso normales, al tiempo que se conserva su utilidad y calidad[150].

Se exige inmediatez a la hora de revelar la información y que la misma se ajuste a los requisitos de accesibilidad aplicables. Así, toda información y notificaciones deberán facilitarse de manera clara y distinguible a más tardar con ocasión de la primera interacción o exposición al sistema (art. 50.5).

4.3.2. Derecho a una explicación

En la versión original del RIA, de 21 de abril de 2021, no se reconocía ningún derecho a ser informado de la existencia de decisiones automatizadas, ni un derecho de acceso a los datos o información del sistema, ni un derecho a recibir una explicación por su mal funcionamiento[151]. Es decir, no se recogían

---

[150] Se precisa con más detalle en el Considerando 134, que «el cumplimiento de esta obligación de transparencia no debe interpretarse como un indicador de que la utilización del sistema de IA o de sus resultados de salida obstaculiza el derecho a la libertad de expresión y el derecho a la libertad de las artes y de las ciencias, garantizados por la Carta, en particular cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, con sujeción a unas garantías adecuadas para los derechos y libertades de terceros […] Además, también conviene prever una obligación de divulgación similar en relación con el texto generado o manipulado por una IA en la medida en que se publique con el fin de informar al público sobre asuntos de interés público, a menos que el contenido generado por la IA haya sido sometido a un proceso de revisión humana o de control editorial y que una persona física o jurídica ejerza la responsabilidad editorial de la publicación del contenido».

[151] NAVAS NAVARRO, S., *Daños ocasionado…, cit.,* p. 73.

en dicho instrumento derechos subjetivos para los afectados por dichas decisiones (usuarios finales), sino que sus previsiones iban dirigidas exclusivamente a los proveedores de tecnología basada en IA respecto a los responsables de su despliegue[152]. Esta carencia fue muy criticada por la doctrina y los organismos internacionales que la calificaron como un punto ciego de la regulación[153] y el Parlamento introdujo una enmienda en la que recogía un *derecho de explicabilidad algorítmica* muy similar al consagrado en el RGPD, pero con la importante diferencia de ser aplicable también a las *decisiones semiautomatizadas*, esto es, a aquellas que adopta una persona con apoyo en los resultados del sistema. De hecho, en el art. 68 *quarter,* titulado «Derecho a explicación de la toma individual de decisiones», se establecía que «toda persona afectada sujeta a una *decisión adoptada por el implementador* [ahora responsable del despliegue] *sobre la base de la información de salida de un sistema de IA de alto riesgo* que produzca *efectos jurídicos o que le afectan significativamente* de una manera que considera que perjudica a su salud, seguridad, derechos fundamentales, bienestar socioeconómico o cualquier otro de sus derechos derivados de las obligaciones establecidas en el presente Reglamento, *tendrá derecho a solicitar al implementador una explicación clara y significativa, de conformidad con el artículo 13, apartado 1, sobre el papel del sistema de IA en el procedimiento de toma de decisiones, los principales parámetros de la decisión adoptada y los datos de entrada correspondientes»*. Añadiendo en el apartado tercero que dicho precepto se aplicaría *sin perjuicio de lo previsto en los artículos 13, 14, 15 y 22 del RGPD*[154].

El objetivo de dicha enmienda era subsanar la laguna de la que adolecía la normativa y conseguir que la transparencia algorítmica no sólo fuera

---

[152] Nos remitimos a lo ya explicado *supra* sobre el art. 13.

[153] Entre los autores que critican dicha omisión del RIA: RIVAS VALLEJO, P., «Herramientas desde el derecho antidiscriminatorio…», *cit.,* p. 542, ÁLVAREZ CUESTA, H. «Inteligencia artificial: derecho de la UE y derecho comparado. La propuesta de una ley sobre IA», *cit.,* o MUÑOZ GARCÍA, C., «Adaptar o reformular la directiva 85/374 sobre responsabilidad por daños causados por productos defectuosos a la inteligencia artificial», accesible en https://diariolaley.laleynext.es/dll/2022/03/01/adaptar-o-reformular-la-directiva-85-374-sobre-responsabilidad-por-danos-causados-por-productos-defectuosos-a-la-inteligencia-artificial. Según la Recomendación de la UNESCO (ap. 38), las personas deben estar plenamente informadas cuando una decisión se basa en algoritmos de IA o se toma a partir de ellos, en particular cuando afecta a su seguridad o derechos humanos y deben tener la oportunidad de solicitar explicaciones e información al actor de la IA o a las instituciones públicas correspondientes. Además, deben poder conocer los motivos por los que se ha tomado una decisión que afecta a sus derechos y libertades y tener la posibilidad de presentar alegaciones a fin de que se revise y enmiende la decisión.

[154] El apartado segundo del art.68 *quarter* establecía excepciones a la aplicación de este derecho de explicación a disponer que «El apartado 1 no se aplicará a la utilización de sistemas de IA para los que el Derecho nacional o de la Unión prevea excepciones o restricciones a la obligación prevista en el apartado 1 en la medida en que dichas excepciones o restricciones respeten la esencia de los derechos y libertades fundamentales y sean una medida necesaria y proporcionada en una sociedad democrática»

dirigida a los operadores de la tecnología sino también a las personas físicas afectadas por las decisiones automatizadas. Tras dicha enmienda, la versión del texto aprobada el 13 de marzo de 2024 por la Eurocámara[155] recogió por primera vez ese *derecho de las personas afectadas a obtener una explicación de las decisiones individuales*, novedad importante respecto de la que no se produjeron modificaciones tras su aprobación definitiva en mayo de ese mismo año.

Dispone ahora literalmente el art. 86 que «Toda persona que se vea afectada por una *decisión que el responsable del despliegue adopte basándose en los resultados de un sistema de IA de alto riesgo* que figure en el anexo III, con excepción de los sistemas enumerados en su punto 2, *y que produzca efectos jurídicos o le afecte considerablemente del mismo modo*, de manera que considere que tiene un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales, *tendrá derecho a obtener del responsable del despliegue explicaciones claras y significativas acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada*». Se consagra expresamente un derecho de explicación *ex post,* central en la ordenación de la IA, que viene a completar el limitado y criticable ámbito del art. 22 RGPD por cuanto se aplica a los sistemas de apoyo a la decisión y no sólo a los que tomen decisiones de forma totalmente automatizada.

Se establecen dos límites a la aplicación de este derecho: el primero, que no se aplicará cuando se utilicen sistemas de IA para los que existan excepciones o restricciones a la obligación prevista en el apartado 1 derivadas del Derecho nacional o de la UE (art. 86.2); el segundo, que solo se aplicará en la medida en que no esté ya previsto en el Derecho de la Unión (art. 86.3), disposición que contiene una clara referencia al RGPD.

Dicho derecho se completa con la obligación de los responsables del despliegue establecida en el nuevo art. 26.11, en virtud de la cual, en el caso de que se estén utilizando sistemas de IA de alto riesgo del anexo III que *tomen decisiones* o *ayuden a tomar decisiones* relacionadas con personas físicas, *deberán informar a dichas personas de que están expuestas a la utilización de tales sistemas*, sin perjuicio de lo dispuesto en el art. 50 sobre obligaciones de transparencia específicas cuando se usen determinados sistemas IA destinados a interactuar con personas físicas[156]. Por tanto, es evidente que los responsables del despliegue asumen un papel fundamental a la hora de garantizar la transparen-

---

[155] P9TA(2024)0138. Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 - C9-0146/2021 - 2021/0106(COD)).

[156] También dispone el art. 26.11 que «En el caso de los sistemas de IA de alto riesgo que se utilicen a los efectos de la aplicación de la ley, se aplicará el artículo 13 de la Directiva (UE) 2016/680».

cia externa hacia las personas afectadas por los sistemas tanto antes (*ex ante*) como después de adoptada la decisión (*ex post*). Respecto a la primera, porque deben informar a las personas físicas de que son objeto de la utilización de un sistema de IA de alto riesgo, información que debe incluir tanto la finalidad prevista como el tipo de decisiones que se toman. Respecto a la segunda, porque deben informales de que gozan de un derecho a obtener una explicación de la decisión y darle cumplimiento.

De todo esto se infiere que, igual que dijimos respecto del RGPD, el RIA también pretende superar los dos niveles de opacidad algorítmica a los que se hizo referencia en el apartado 2. Por un lado, garantizando la transparencia *sobre el uso* de dichos sistemas para que los ciudadanos sean conscientes de que es una IA la que decide sobre su persona; por otro lado, garantizando la transparencia *sobre el contenido* y justificación de la decisión final.

Por lo que se refiere al contenido concreto de la explicación que debe ofrecer el responsable al afectado por la decisión, dos son los extremos sobre los que ha de informarle: en primer lugar, del papel que el sistema de IA ha tenido en el proceso de toma de decisiones; y en segundo lugar, de los principales elementos de la decisión adoptada. Deberá hacerlo de forma clara y significativa, es decir, de forma suficiente para que el afectado pueda ejercer sus derechos.

Dar explicaciones sobre el *papel que ha jugado el sistema* a la hora de tomar la decisión implicará informar al interesado sobre si se trata de una decisión plena o semiautomatizada, es decir, si el resultado se ejecutó de forma totalmente automática y sin intervención humana, o si se utilizó como mero apoyo de la decisión adoptada finalmente por una persona. Además, en este último caso tendrá que indicar también si el resultado del sistema fue *determinante o no* para la toma de la decisión final y en qué medida lo fue[157].

De hecho, la cuestión de hasta qué punto la salida o resultado del sistema de IA es *determinante* para la decisión finalmente adoptada por el usuario, ya no es solo relevante a los efectos de aplicar las garantías del art. 22 RGPD, sino que ahora también es decisiva para determinar si un sistema de IA es considerado de alto riesgo y, en consecuencia, para la aplicación de todas las garantías previstas en el RIA[158]. Es decir, los sistemas de IA independientes se clasifican

---

[157] Sobre el carácter «determinante» o no del resultado del sistema para la toma de la decisión final nos remitimos a las apreciaciones realizadas al comentar *supra* la STJUE de 7 de diciembre de 2023.

[158] Así lo aprecia con acierto Cotino Hueso, «La primera sentencia del Tribunal de Justicia de la Unión Europea…», cit., para quien, la relevancia del resultado del sistema de IA en la decisión finalmente adoptada ha pasado a ser en el RIA un elemento esencial para determinar si el sistema es de alto riesgo.

como de alto riesgo si, a la luz de su *finalidad prevista*, presentan un *alto riesgo de ser perjudiciales para la salud y la seguridad o los derechos fundamentales de las personas*, teniendo en cuenta tanto la *gravedad* del posible perjuicio como la *probabilidad* de que se produzca, y se utilizan en una serie de ámbitos especificados en el Reglamento. No obstante, tal como se explica en el Considerando 53, pueden existir casos en los que, a pesar de utilizarse en esos ámbitos predefinidos, los sistemas no entrañen un riesgo considerable de causar perjuicios a intereses jurídicos protegidos *si no influyen sustancialmente en la toma de decisiones*. Se entenderá que no influyen sustancialmente en la toma de decisiones cuando no afecten al fondo, ni por consiguiente al resultado, de la toma de decisiones humana o automatizada. Las condiciones (una o varias) para que pueda considerarse que un sistema no plantea dicho riesgo al no influir sustancialmente en el resultado de la toma de decisiones, se recogen en el art. 6.3. Su concurrencia implicará que esos sistemas no queden sometidos a los requisitos que el RIA impone a los sistemas de alto riesgo.

Respecto a los *principales elementos de la decisión adoptada*, creemos que podría aplicarse aquí lo explicado en el apartado 4.2.2 sobre el contenido del derecho de información reforzado en el RGPD. Así, habría que informar, entre otras cosas, de los factores utilizados por el algoritmo para tomar la decisión o la elaboración del perfil, la ponderación relativa de cada variable en el modelo para la toma de la decisión y cualquier cambio de estos parámetros que modifique el comportamiento del algoritmo, las reglas e instrucciones utilizadas por el algoritmo, así como los datos utilizados por el sistema respetando en todo caso el RGPD.

A nuestro modo de ver, debe aplaudirse que el RIA haya acabado consagrando este derecho a favor de los potenciales afectados por las decisiones de la IA en la medida en que contribuirá a que comprendan cómo se adoptan y cómo pueden afectarles. A este fin contribuye también la exigencia del RIA de educar y formar en competencias de IA a todos los que operan con ella. En su art. 4, titulado «Alfabetización en materia de inteligencia artificial» se establece que «*Los proveedores y responsables del despliegue de sistemas de IA adoptarán medidas para garantizar que, en la mayor medida posible, su personal*[159] *y demás personas que se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA tengan un nivel suficiente de alfabetización en materia de IA, teniendo en cuenta*

[159] Asimismo, los responsables del despliegue deben garantizar que las personas encargadas de poner en práctica las instrucciones de uso y la supervisión humana establecidas en el presente Reglamento tengan las competencias necesarias, en particular un nivel adecuado de alfabetización, formación y autoridad en materia de IA para desempeñar adecuadamente dichas tareas. Dichas obligaciones deben entenderse sin perjuicio de otras obligaciones que tenga el responsable del despliegue en relación con los sistemas de IA de alto riesgo con arreglo al Derecho nacional o de la Unión (Considerando 91).

sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso de los sistemas de IA y las personas o los grupos de personas en que se utilizarán dichos sistemas»[160].

En el Considerando 20 se explica con más detalle que, con el fin de aprovechar las ventajas de estos sistemas y proteger al mismo tiempo los derechos fundamentales, la salud y la seguridad, y de posibilitar el control democrático, la alfabetización en IA debe dotar a proveedores, responsables del despliegue y personas afectadas, de los conceptos necesarios para *tomar decisiones con conocimiento de causa en relación con los sistemas de IA*[161]. Por lo que se refiere concretamente a estas últimas, se les debe dotar de los *conocimientos necesarios para comprender el modo en que las decisiones adoptadas con la ayuda de la IA tendrán repercusiones para ellas.*

Dicha formación será importante, además, para que cualquier persona que deba tomar decisiones basadas en resultados algorítmicos pueda hacer frente de forma eficaz al peligroso sesgo de la automatización[162]. Asimismo, estos conocimientos previos permitirán, posteriormente, implementar otras medidas desde el ámbito de la tecnología para abordar el sesgo de estos sistemas, y en particular, el sesgo en los datos que los alimentan, el cual se erige como principal causa de la discriminación algorítmica. En definitiva, con esta alfabetización en IA se pretende, por un lado, proporcionar a todos los agen-

---

[160] El art.3, ap. 56, define la «alfabetización en materia de inteligencia artificial» como «las capacidades, los conocimientos y la comprensión que permiten a los proveedores, responsables del despliegue y demás personas afectadas, teniendo en cuenta sus respectivos derechos y obligaciones en el contexto del presente Reglamento, llevar a cabo un despliegue informado de los sistemas de IA y tomar conciencia de las oportunidades y los riesgos que plantea la inteligencia artificial, así como de los perjuicios que puede causar». Además, se establece que el Comité Europeo de Inteligencia Artificial deberá apoyar a la Comisión para promover las herramientas de alfabetización en IA, la sensibilización pública y la comprensión de los beneficios, riesgos, salvaguardias, derechos y obligaciones en relación con el uso de sistemas de IA (vid., art. 66 «Funciones del Comité», concretamente, el ap. f), y que tanto la Comisión como los Estados miembros deberán facilitar la elaboración de códigos de conducta voluntarios para promover dicha alfabetización entre las personas que se ocupan del desarrollo, manejo y uso de la IA (vid., art. 95.c).

[161] Estos conceptos pueden variar en función del contexto pertinente e incluir el entendimiento de la correcta aplicación de los elementos técnicos durante la fase de desarrollo del sistema de IA, las medidas que deben aplicarse durante su uso, las formas adecuadas de interpretar la información de salida del sistema de IA y, en el caso de las personas afectadas, los conocimientos *necesarios para comprender el modo en que las decisiones adoptadas con la ayuda de la IA tendrán repercusiones para ellas.*

[162] Para SORIANO ARNANZ, A./SIMÓ SOLER, E. (2021), «Machine learning y Derecho: aprendiendo la (des)igualdad», en *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*, S. Barona Vilar (ed.), Valencia, p. 203, la ventaja de que jueces y magistrados estén debidamente formados en perspectiva de género no se limita a que sus pronunciamientos carezcan de prejuicios negativos, sino que cumple una función preventiva y de detección de argumentos estereotipados por parte de los empresarios que utilizan sistemas de IA para la toma de decisiones, como de los propios programadores que introducen las instrucciones, variables y datos.

tes de la cadena de valor de la IA los conocimientos necesarios para garantizar el cumplimiento y la correcta ejecución del RIA, y por otro lado, que tanto la sociedad en general como sobre todo los que se encargan de diseñar, desarrollar, desplegar o usar estos sistemas, tomen conciencia de sus características, forma de funcionamiento, ventajas y potenciales riesgos. Todo ello coadyuvará a consolidar el camino hacia una IA fiable.

5. CONSIDERACIONES FINALES SOBRE LA NECESIDAD DE TRANSPARENCIA Y EXPLICABILIDAD PARA DETECTAR Y DEMOSTRAR LA DISCRIMINACIÓN ALGORÍTMICA

La transparencia ha sido una cuestión prioritaria a la que se ha dedicado especial atención por las instituciones y la doctrina desde los inicios del tratamiento regulatorio de la IA. De hecho, se ha erigido en un principio básico de muchos marcos normativos, éticos y de seguridad vigentes y proyectados. Es lógico, por ello, que también sea uno de los pilares básicos sobre los que se asienta el reciente Reglamento europeo que regula la materia y con el que se pretende fomentar la confianza en los sistemas inteligentes y mitigar riesgos como el sesgo y la discriminación injusta al que se ha hecho especial referencia en este trabajo.

Si bien es cierto que la transparencia no puede impedir que los sistemas de IA lleguen a resultados errados y/o discriminatorios, ni garantizar su equidad y justicia, es fundamental para descubrir desviaciones que pasarían desapercibidas si no se exigiera tanto en el diseño como en el despliegue y uso de los sistemas de IA. Dicha exigencia debe operar en todo el ciclo de vida de los sistemas y ha de ser tanto interna como externa.

Es más, no solo ha de abogarse porque los sistemas sean transparentes, sino también, y sobre todo, porque sus resultados sean explicables. Mientras la transparencia tiene que ver con el proceso seguido para llegar a un resultado, y pretende dar a conocer y entender cómo se procesa el dato que se introduce en el sistema, la explicabilidad tiene que ver con el resultado de ese proceso y desciende más al detalle[163]. Por tanto, de lo que se trataría con la explicabilidad sería de que el afectado por la decisión pudiera conocer los motivos por los que se ha tomado una decisión que afecta a sus derechos y libertades y tuviera la posibilidad de impugnarlas. No sería suficiente con dar una explicación genérica sobre cómo funciona el proceso de toma de decisión, sino que

[163] Así lo afirma el Director del Centro Europeo para la Transparencia Algorítmica (ECAT), Alberto Pena Fernández, en Diario La Ley, 8 de marzo de 2024.

habría que indicar los datos específicos de la persona en base a los cuales se ha tomado una decisión o se ha hecho un perfilado. En definitiva, no solo habría que informar de forma clara y sencilla de la denominada «lógica computacional» del sistema, sino también de su «lógica decisional». Sólo así podrán salvaguardarse derechos fundamentales como la tutela judicial efectiva, el derecho a la defensa o la igualdad y no discriminación.

En este sentido, es elogiable que el RIA haya acabado acogiendo un *derecho a la explicación* que viene a completar tanto las previsiones sobre transparencia y comunicación de información entre los diversos sujetos de la cadena de valor (principalmente, entre proveedores y responsables del despliegue) que ya preveía su versión inicial, como el insuficiente y criticado alcance del art. 22 RGPD que suponía una importante traba en la lucha contra la opacidad de los sistemas de IA.

Asimismo, cuando las decisiones se basan en datos personales, el control sobre nuestra información personal ha de implicar la posibilidad de ejercer, como mínimo, una serie de derechos como el de ser informados sobre el razonamiento subyacente al procesamiento de datos realizado por los algoritmos, el derecho a oponerse a dicho procesamiento y el derecho a recurrirlo. Ello no sólo cuando se trate de decisiones plenamente automatizadas sino también semiautomatizadas.

En definitiva, tanto la transparencia como la explicabilidad de los sistemas de IA que toman decisiones basadas en datos personales y con efectos jurídicos en personas concretas son cruciales para que la ciudadanía confíe en ellos. Deben articularse medidas para hacer frente a la opacidad algorítmica tanto desde el ámbito jurídico, como desde el terreno ético y por supuesto tecnológico, tratando de encontrar un equilibrio adecuado entre las exigencias requeridas para proteger la salud, seguridad y derechos fundamentales de las personas y los intereses empresariales. Sin duda, tanto la aprobación definitiva del RIA, como la nueva interpretación del TJUE sobre el ámbito de aplicación del art. 22 RGPD, suponen pasos en firme para conseguirlo.

BIBLIOGRAFÍA

ALAMEDA CASTILLO, M.T., «Reclutamiento tecnológico. Sobre algoritmos y acceso al empleo», *Temas laborales: Revista andaluza de trabajo y bienestar social*, Nº 159, 2021, pp. 11-52.

ALCOLEA AZCÁRRAGA, C., «La responsabilidad patrimonial de la Administración y el uso de algoritmos», *Revista General de Derecho Administrativo*, nº 59, enero 2022. Recuperado de https://laadministracionaldia.inap.es/noticia.asp?id=1512629

ÁLVAREZ CUESTA, H. «Inteligencia artificial: derecho de la UE y derecho comparado. La propuesta de una ley sobre IA», en P. Rivas Vallejo (dir.), *Discriminación algorítmica en el ámbito laboral*, Navarra 2022, pp. 379-414.

ARCOS VARGAS, M., «La no discriminación en el Derecho derivado de la Unión Europea» en J.M. MORALES ORTEGA (dir.), *Realidad social y discriminación. Estudios sobre diversidad e inclusión laboral*, Murcia, 2022, pp. 17-41.

ATIENZA NAVARRO, M.L, *Daños causados por inteligencia artificial y responsabilidad civil*, 2022.

BARRIOS ANDRÉS, M., *Manual de Derecho Digital*, Valencia, 2020.

BAZ LOMBA, C. (2021). Los algoritmos y la toma de decisiones administrativas. Especial referencia a la transparencia. *Revista CEFLegal*, 243, 119-160.

BELTRÁN DE HEREDIA, I., «Automatización y obsolescencia humana», en *Retos jurídicos de la inteligencia artificial*, coord. por Agustí Cerrillo i Martínez y Miquel Peguera Poch, 2020, pp. 113-124.

BIBAL A./LOGNOUL, M./DE STREEL, A./FRÉNAY, B., «Legal requirements on explainability in machine learning», *Artificial Intelligence & Law*, 29(2), 2021, pp. 149-169.

CASADESÚS RIPOLL, P., *La responsabilidad civil por el uso discriminatorio de los datos personales a través de la inteligencia artificia*, Granada, 2022.

CASTELLANOS CLARAMUNT, J., «Derecho e inteligencia artificial: atención especial a los sesgos, la privacidad y la protección de datos», accesible en https://idpbarcelona.net/derecho-e-inteligencia-artificial-atencion-especial-a-los-sesgos-la-privacidad-y-la-proteccion-de-datos/

CASTILLA, K., *Cuatro ángulos de análisis de la igualdad y la no discriminación en la inteligencia artificial*, Barcelona, 2022, https://www.idhc.org/arxius/recerca/Inteligencia-artificial-vF.pdf

CASTILLO OCARANZA, C., «Discriminación algorítmica en el ámbito laboral», en P. Rivas Vallejo (dir.), *Discriminación algorítmica en el ámbito laboral*, Navarra 2022, pp. 71-78.

CASTILLO PARRILLA, J.A., «Sentencia del Tribunal Ordinario de Bolonia de 31 de diciembre de 2020 (Caso Deliveroo) ¿Discriminación algorítmica o discriminación a través de un algoritmo?», *Derecho Digital e Innovación. Digital Law and Innovation Review*, n.7 (octubre-diciembre), 2020.

CONSEJO DE EUROPA, *Manual de legislación europea contra la discriminación (edición 2018)*, Luxemburgo, 2019.

COTINO HUESO, L., «La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial» *Diario LA LEY*, Nº 80, Sección Ciberderecho, 17 de Enero de 2024.

— «Caso Bosco, a la tercera tampoco va la vencida. Mal camino en el acceso los algoritmos públicos», Diario LA LEY, Nº 84, Sección Ciberderecho, 17 de Mayo de 2024.

— «Transparencia de la inteligencia artificial pública: marco legal, desafíos y propuestas», *Actualidad Administrativa, Nº I, Sección Actualidad,* Diciembre 2023.

DE LA NUEZ SÁNCHEZ CASCADO, E., «Algoritmos y transparencia», disponible en https://www.hayderecho.com/2020/02/18/algoritmos-y-transparencia-2/.

EGUÍLUZ CASTAÑEIRA. J.A., «Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales», *Estudios de Deusto: revista de Derecho Público*, vol. 68, n. 2, 2020, pp. 325-367.

EUBANKS, V., *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*, St MARTÍN's Press, 2018.

EVANGELIO LLORCA, R., «Causalidad y responsabilidad civil por daños ocasionados por sistemas de inteligencia artificial: las presunciones de causalidad en las propuestas normativas de la UE», en *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial,* coord. por N. Álvarez Lata; J.M. Busto Lago (pr.), 2024, pp. 549-619.

FERNÁNDEZ DE LA MORENA, B., *Discriminación algorítmica. Estudio del sesgo en arquitecturas de aprendizaje profundo*, Madrid, 2019.

FERNÁNDEZ ROZAS, J.C., «La Ley de Inteligencia Artificial de la Unión Europea: un modelo para innovaciones radicales, responsables y transparentes basadas en el riesgo», LA LEY Unión Europea, Nº 124, Sección Estudios, Abril 2024, LA LEY 15196/2024.

GASCÓN MARCÉN, A., «Derechos humanos e inteligencia artificial», en *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Pérez Miras (dir.), Teruel Lozano (dir.), Raffiotta (dir.), Pia Iadicco (dir.), Vol. 5, 2020 (Retos en el siglo XXI / coord. por Silvia Romboli), pp. 335-350.

GERARDS, J./ XENIDIS, R., *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law* (Special report European network of legal experts in gender equality and non-discrimination), European Commission, 2020.

GIL MEMBRADO, C., «Daños producidos por la IA: la opacidad del algoritmo y el efecto de caja negra», en N. Álvarez Lata/J.M. Busto Lago, *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*, 2024, pp. 501-547.

GINÈS i FABRELLAS, A., «Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica», en RIVAS VALLEJO, P. (dir), *Discriminación algorítmica en el ámbito laboral*, Navarra, 2022, pp. 295-332.

GONZÁLEZ VALVERDE, A., «Responsabilidad por el potencial de sesgo discriminatorio de los algoritmos de los productos de Inteligencia Artificial. A propósito de la Algorithmic Accountability Act of 2019 (s.1108)». En Ataz López y Cobacho Gómez, *Cuestiones clásicas y actuales del Derecho de daños: Estudios en homenaje al profesor Dr. Roca Guillamón,* Vol. 2, 2021, pp. 1225-1273.

GOODMAN., B/FLAXMAN, S., «European Union regulations on algorithmic decision-making and a "right to explanation"», presented at ICML Workshop on Human Interpretability in Machine Learning, New York, 2016, https://arxiv.org/abs/1606.08813v3.

HERRERA DE LAS HERAS, R., «Protección de datos e inteligencia artificial», en Cruz Blanca/Lledó Benito (coords.), *La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0*, 2021, pp. 644-669.

HUERGO LORA, A., «Una aproximación a los algoritmos desde el derecho administrativo», en A. HUERGO LORA (dir.) y G. DÍAZ GONZÁLEZ (coord.), *La regulación de los algoritmos*, Navarra, 2020, pp. 23-87.

— «El proyecto de Reglamento sobre la Inteligencia Artificial», *Almacén del Derecho,* 17 de abril de 2021.

— «Por qué aciertan las sentencias sobre el 'algoritmo' del bono social eléctrico», *Almacén del Derecho*, 10 de mayo de 2024.

LAZCOZ MORATINOS, G., «Análisis jurídico de la toma de decisiones algorítmica en la asistencia sanitaria», en *La regulación de los algoritmos*, coord. por Gustavo Manuel Díaz González; Alejandro José Huergo Lora (dir.), 2020, pp. 283-297.

LÓPEZ-TARRUELLA MARTÍNEZ, A., *Propiedad intelectual e innovación basada en los datos*, Navarra, 2021.

LLORENS ESPADA, J., «Responsabilidades civiles por discriminación por razón de género cuando medie un sistema de inteligencia artificial», en Rivas Vallejo (dir.), *Discriminación algorítmica en el ámbito laboral*, Navarra 2022, pp. 573-602.

MANHEIM, K. M., & KAPLAN, L., «Artificial Intelligence: Risks to Privacy and Democracy», Yale Journal of Law and Technology 106, vol. 21, 2019, pp. 106-189.

MEDINA GUERRERO, M., «El derecho a conocer los algoritmos utilizados en la toma de decisiones. Aproximación desde la perspectiva del derecho fundamental a la protección de datos personales», *Teoría y realidad constitucional*, nº 49, 2022, pp. 141-171.

MERCADER UGUINA, J.R. «Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del big data», *Labos: Revista de Derecho del Trabajo y Protección Social*, vol. 2, n. 2, 2021, pp. 4-10.

MINISTERIO DE TRABAJO Y ECONOMÍA SOCIAL, *Información algorítmica en el ámbito laboral. Guía Práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral,* mayo 2022.

MUÑOZ GUTIÉRREZ, C., «La discriminación en una sociedad automatizada: Contribuciones desde América Latina», *Revista chilena de derecho y tecnología,* 10(1), 2021, pp. 271–307.

MUÑOZ GARCÍA, C., «Adaptar o reformular la directiva 85/374 sobre responsabilidad por daños causados por productos defectuosos a la inteligencia artificial», *Revista Crítica de Derecho Inmobiliario,* Año 98, nº 793, 2022, pp. 2886-2908.

NAVAS NAVARRO, S., *Daños ocasionados por sistemas de inteligencia artificial. Especial atención a su futura regulación,* Granada, 2022.

— «La perspectiva de género en la inteligencia artificial», *Diario La Ley,* nº 48, Sección Ciberderecho, 8 de marzo de 2021.

— «Sistemas expertos basados en inteligencia artificial y responsabilidad civil. Algunas cuestiones controvertidas», *Diario La Ley*, 13 de Diciembre de 2019, LA LEY 14980/2019.

NÚÑEZ SEOANE, J., «El derecho a la información y acceso al funcionamiento de los algoritmos que tratan datos personales», en *La regulación de los algoritmos*, coord. por Gustavo Manuel Díaz González y Alejandro José Huergo Lora (dir.) 2020, pp. 299-315.

O'NEIL, C., *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* Madrid, 2017.

OTTOLIA, A., *Derecho, Biga Bata e Inteligencia Artificial*, Valencia, 2018.

PALMA ORTIGOSA, A., «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», *Revista General de Derecho Administrativo*, n. 50, 2019.

REBOLLO DELGADO, L. *Inteligencia artificial y derechos fundamentales*, Madrid, 2023.

RIVAS VALLEJO, P., «Sesgos de automatización y discriminación algorítmica», en *Discriminación algorítmica en el ámbito laboral*, Pilar Rivas Vallejo (dir.), 2022, pp. 31-68.

— «Análisis desde el derecho antidiscriminatorio», en P. Rivas Vallejo (dir.*) Discriminación algorítmica en el ámbito laboral*, Navarra 2022, pp.415-476.

— «Herramientas desde el derecho antidiscriminatorio y la protección frente a la igualdad y a la no discriminación», en P. Rivas Vallejo (dir.*) Discriminación algorítmica en el ámbito laboral*, Navarra 2022, pp. 539-572.

RUBÍ PUIG, A., «Elaboración de perfiles y personalización de ofertas y precios en la contratación con consumidores», *Revista De Educación Y Derecho*, 24, 2021.

SÁEZ LARA, C., «El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación», *Temas laborales: Revista andaluza de trabajo y bienestar social*, Nº 155, 2020, pp. 41-60.

SÁNCHEZ CAPARRÓS, M. «Prevenir y controlar la discriminación algorítmica», 2022. https://www.researchgate.net/publication/358207305_Prevenir_y_controlar_la_discriminacion_algoritmica

SELBST, A. D./POWLES, J., «Meaningful information and the right to explanation», *International Data Privacy Law*, 2017, Vol. 7, n.º 4, pp. 233 y ss.

SOLAR CAYÓN, J.I., «Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos», en *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho,* J.I. Solar Cayón (ed.), Alcalá de Henares, 2020, pp. 125-172.

SORIANO ARNANZ, A. «Decisiones automatizadas y discriminación: aproximación y propuestas generales», *Revista General de Derecho Administrativo*, n. 56, 2021.

— «La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de Inteligencia Artificial», en *Nuevas normatividades: inteligencia artificial, derecho y género* (coord. por P.R. Bonorino Ramírez / R. Fernández Acevedo / P. Valcárcel Fernández), Navarra, 2022, pp. 63-88.

SORIANO ARNANZ, A./SIMÓ SOLER, E., «Machine learning y Derecho: aprendiendo la (des)igualdad», en S. Barona Vilar (ed.) *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*, Valencia, 2021, pp. 183-207.

SUBERBIOLA GARBIZU, I., «Inteligencia artificial, opacidad y motivación de las decisiones administrativas automatizadas», en *La protección de los derechos fundamentales en el ámbito tributario*, coord. por Antonio Vázquez del Rey Villanueva e Irune Suberbiola Garbizu y dir., por Isaac Merino Jara, 2021, pp. 291-328.

TOLOSA, P./DIBO, C., «Inteligencia artificial, discriminación por género y Derecho: viejos problemas, nuevos desafíos», en C. Danesi (dir.) *Inteligencia Artificial, tecnologías emergentes y Derecho,*2021, pp. 155-190.

VESTRI, G., «La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativ», *Revista Aragonesa de Administración Pública*, nº 56, 2021, pp. 368-398.

WACHTER, S., MITTELSTADT, B., y FLORIDI, L. «Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation», *International Data Privacy Law*, Vol. 7, n. 2, 2017, pp. 1-47.

La inteligencia artificial tiene el potencial de transformar productos, servicios y procedimientos en multitud de sectores económicos y en relación con muchos ámbitos de la sociedad. Sin embargo, también puede generar un sinfín de riesgos que, de producir daños, habrán de ser reparados. La Unión Europea no ha sido ajena a estos riesgos, y por ello ha pretendido y sigue pretendiendo crear un marco jurídico protector. Dentro de este contexto, se sitúa la aprobación del Reglamento (UE) 1689 del Parlamento y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial -RIA-, como sendas Propuestas de Directiva, de inminente aprobación, sobre responsabilidad civil de productos defectuosos y sobre responsabilidad civil por daños causados por la inteligencia artificial. Partiendo de tales postulados, en la presente obra se han seleccionado aquellos sectores donde, por su mayor proyección, novedad o complejidad, merece ser analizada la interrelación entre la tecnología de la inteligencia artificial y el Derecho de daños. Para ello, se ha podido contar con un elenco de especialistas en el sector, que sin duda hace de la obra resultante una aportación doctrinal de indudable utilidad.

Con carácter particular, entre los sectores seleccionados, destaca por su trascendencia, el de la salud digital, donde problemáticas relacionadas con sistemas inteligentes para la prevención de enfermedades, ya sea a iniciativa del profesional de la medicina, o al margen de él -uso de werables y servicios digitales-, o por infracciones de los datos personales de salud, pueden determinar, si bien a través de distintos cauces normativos, posibles vías de reclamación indemnizatoria.

En el campo quirúrgico, la "cirugía 4.0", que integra la cirugía robótica y personalizada, por su creciente implantación, ha merecido una especial consideración en la obra.

Se efectúan igualmente amplías consideraciones acerca de la transparencia y explicabilidad para prevenir la discriminación algorítmica en el uso de los sistemas de inteligencia artificial.

Dentro de los sectores con mayor implementación de las tecnologías de inteligencia ha sido objeto de consideración así mismo el uso de vehículos autónomos, incluida su problemática en la vertiente del Derecho internacional privado.

Situados en el marco normativo que proporciona el Reglamento de Inteligencia artificial -RIA- se efectúan correspondientes análisis acerca de la categorización del riesgo que el mismo contempla, y donde se observa un régimen jurídico tendente a salvaguardar los riesgos más graves por el empleo de los sistemas de inteligencia artificial; en particular, en la salud, seguridad y derechos consagrados en la Carta Europea de Derechos Fundamentales. De igual forma las implicaciones jurídicas que despliega la inteligencia artificial generativa por infracciones normativas del Derecho de protección de datos personales. Se incluyen también los rasgos que deben estar presentes en el seguro de responsabilidad civil profesional de los operadores de inteligencia artificial, a partir de las previsiones normativas del referido Reglamento de Inteligencia Artificial.

9 788410 707085