ESTUDIOS

DERECHO E INTELIGENCIA ARTIFICIAL

AVANZANDO EN SU REGULACIÓN

INMACULADA HERBOSA MARTÍNEZ
DAVID FERNÁNDEZ DE RETANA GOROSTIZAGOIZA
DIRECTORES

INCLUYE LIBRO ELECTRÓNICO

ARANZADI

© **Inmaculada Herbosa Martínez y David Fernández de Retana Gorostizagoiza (Dirs.) y autores**, 2025
© **ARANZADI LA LEY, S.A.U.**

ARANZADI LA LEY, S.A.U.
C/ Collado Mediano, 9
28231 Las Rozas (Madrid)
www.aranzadilaley.es
Atención al cliente: https://areacliente.aranzadilaley.es/

Primera edición: Marzo 2025

Depósito Legal: M-7454-2025
ISBN versión impresa con complemento electrónico: 978-84-1162-707-8
ISBN versión electrónica: 978-84-1162-709-2

Para la publicación de este libro/capítulo se ha contado con la ayuda del Departamento de Educación del Gobierno Vasco destinada a apoyar las actividades de los grupos de investigación del Sistema Universitario Vasco (Ref. IT1472-22).

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

© **ARANZADI LA LEY, S.A.U.** Todos los derechos reservados. A los efectos del art. 32 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba la Ley de Propiedad Intelectual, ARANZADI LA LEY, S.A.U., se opone expresamente a cualquier utilización del contenido de esta publicación sin su expresa autorización, lo cual incluye especialmente cualquier reproducción, modificación, registro, copia, explotación, distribución, comunicación, transmisión, envío, reutilización, publicación, tratamiento o cualquier otra utilización total o parcial en cualquier modo, medio o formato de esta publicación.

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la Ley. Diríjase a **Cedro** (Centro Español de Derechos Reprográficos, **www.cedro.org**) si necesita fotocopiar o escanear algún fragmento de esta obra.

El editor y los autores no asumirán ningún tipo de responsabilidad que pueda derivarse frente a terceros como consecuencia de la utilización total o parcial de cualquier modo y en cualquier medio o formato de esta publicación (reproducción, modificación, registro, copia, explotación, distribución, comunicación pública, transformación, publicación, reutilización, etc.) que no haya sido expresa y previamente autorizada.

El editor y los autores no aceptarán responsabilidades por las posibles consecuencias ocasionadas a las personas naturales o jurídicas que actúen o dejen de actuar como resultado de alguna información contenida en esta publicación.

ARANZADI LA LEY no será responsable de las opiniones vertidas por los autores de los contenidos, así como en foros, chats, u cualesquiera otras herramientas de participación. Igualmente, ARANZADI LA LEY se exime de las posibles vulneraciones de derechos de propiedad intelectual y que sean imputables a dichos autores.

ARANZADI LA LEY queda eximida de cualquier responsabilidad por los daños y perjuicios de toda naturaleza que puedan deberse a la falta de veracidad, exactitud, exhaustividad y/o actualidad de los contenidos transmitidos, difundidos, almacenados, puestos a disposición o recibidos, obtenidos o a los que se haya accedido a través de sus PRODUCTOS. Ni tampoco por los Contenidos prestados u ofertados por terceras personas o entidades.

ARANZADI LA LEY se reserva el derecho de eliminación de aquellos contenidos que resulten inveraces, inexactos y contrarios a la ley, la moral, el orden público y las buenas costumbres.

Nota de la Editorial: El texto de las resoluciones judiciales contenido en las publicaciones y productos de **ARANZADI LA LEY, S.A.U.**, es suministrado por el Centro de Documentación Judicial del Consejo General del Poder Judicial (Cendoj), excepto aquellas que puntualmente nos han sido proporcionadas por parte de los gabinetes de comunicación de los órganos judiciales colegiados. El Cendoj es el único organismo legalmente facultado para la recopilación de dichas resoluciones. El tratamiento de los datos de carácter personal contenidos en dichas resoluciones es realizado directamente por el citado organismo, desde julio de 2003, con sus propios criterios en cumplimiento de la normativa vigente sobre el particular, siendo por tanto de su exclusiva responsabilidad cualquier error o incidencia en esta materia.

Índice General

Página

PRÓLOGO 7

PARTE PRIMERA
MARCO GENERAL DE LA INTELIGENCIA ARTIFICIAL

Su incidencia en sectores relevantes
STATHIS BANAKAS, JOSU ANDONI EGUÍLUZ CASTAÑEIRA, CARLOS FERNÁNDEZ HERNÁNDEZ, AITZIBER EMALDI CIRION, MARÍA BURZACO SAMPER, OSCAR MONJE BALMASEDA 25

WHO DECIDES WHO DECIDES? ALGORITHM OR HUMAN? RIGHT AND WRONG IN THE AGE OF ARTIFICIAL INTELLIGENCE
STATHIS BANAKAS 25

1. Two big issues 25
1.1. Regulating algorithms 25
1.2. Algorithms in dispute resolution 26
2. Justice human and non-human 28
3. Algorithms, justice, and the tautological nature of mathematics 29
4. A theory of relativity 31
5. Algorithms, justice and politics 35
6. Elegant justice: ars boni et aequi 36
7. The fight for rights: emotional justice 37
Conclusions 39

Página

REGLAMENTO IA: ASPECTOS ESENCIALES DE UNA NORMA PIONERA
JOSU ANDONI EGUÍLUZ CASTAÑEIRA 41

1. Introducción 43
2. Contexto y relevancia del RIA 46
3. Cronología del RIA 49
4. Objetivo y ámbito de aplicación 54
5. Definiciones esenciales 57
5.1. Conceptos tecnológicos 58
5.1.1. Sistema de IA 58
5.1.2. Sistema de IA de uso general 60
5.1.3. Modelos de IA de uso general 62
5.1.4. Diferencia entre sistema de IA y modelo de IA . . . 64
5.2. Actores clave 65
5.2.1. Proveedor 65
5.2.2. Responsable de despliegue 66
5.2.3. Transición de operador a proveedor posterior . . . 66
6. Enfoque basado en el riesgo 67
6.1. Sistemas de IA prohibidos 67
6.2. Sistemas de alto riesgo 71
6.2.1. Categorización de los sistemas de alto riesgo 71
6.2.2. Ciertas obligaciones de los sistemas de alto riesgo 74
6.3. Obligaciones de transparencia 77
6.4. Modelos de IA de uso general 80
7. Gobernanza y ejecución 82
7.1. Gobernanza europea 83
7.2. Gobernanza nacional 84
8. Vías de recurso y sanciones 86

Página

8.1. Vías de recurso 86
8.2. Sanciones 89
Conclusiones 91
Bibliografía 93

LA NUEVA REGULACIÓN EUROPEA DE LOS DATOS Y LA INTELIGENCIA ARTIFICIAL
CARLOS FERNÁNDEZ HERNÁNDEZ 109

1. Introducción. 111
2. De la estrategia para el mercado único digital de 2015 a la estrategia para la década digital 2030 112
3. La estrategia digital europea de 2020 116
3.1. Objetivos de la Estrategia 117
3.2. Las acciones clave de la Estrategia 118
3.3. La actualización de la Estrategia de 2020 124
A) La Brújula Digital 2030 124
B) El programa estratégico de la Década Digital 2030 125
4. La estrategia europea del dato 129
4.1. El concepto europeo de dato 129
4.2. Características de los datos 130
4.3. Problemas con los datos 130
4.4. Las consecuencias de los problemas 132
4.5. Los pilares de la estrategia 134
A) Los espacios europeos de datos 134
B) La Nube Europea de la Ciencia Abierta 136
4.6. El marco europeo de los datos 136
4.7. El tratamiento de datos personales. 137
5. La Directiva sobre datos abiertos de 2019 138
5.1. Objeto 140

Página

5.2. Datos de alto valor 141
5.3. Formatos abiertos 142
6. El Reglamento sobre gobernanza de datos de 2022 142
6.1. Premisas 143
6.2. Objetivo 143
6.3. Servicios de intermediación de datos 144
6.4. Altruismo de datos 146
6.5. Flujos internacionales de datos 146
7. El Reglamento de datos de 2023 147
7.1. Objeto 147
7.2. Ámbito de aplicación 148
7.3. Intercambio de datos de empresa a consumidor 149
7.4. Obligaciones del vendedor de un producto conectado 149
7.5. Obligaciones de los prestadores de servicios relacionados 150
7.6. Derecho del usuario a compartir datos con terceros 151
7.7. Cláusulas contractuales 151
8. La nueva regulación europea de la IA 153
8.1. La obligación de los sistemas de IA de alto riesgo de utilizar datos de calidad y un sistema de gobernanza de datos 154
8.2. Criterios de calidad que deben cumplir los datos de entrenamiento, validación y prueba de los sistemas de IA 155
8.3. Noción de sesgo en los datos 155
8.4. Tratamiento excepcional de datos de categoría especial para prevenir y corregir sesgos 156

PRINCIPALES DERECHOS AFECTADOS EN LA ERA DIGITAL Y DE LA INTELIGENCIA ARTIFICIAL EN EUROPA
AITZIBER EMALDI CIRIÓN 159

1. Introducción 160
2. La seguridad en Europa en la era digital 161

Página

3. Líneas de actuación de los poderes públicos frente a los ciberataques 165

3.1. Sujetos activos de los ciberataques 166

3.2. Ámbitos de actuación de los poderes públicos para combatir los ciberataques 169

4. Vulneración de derechos constitucionales derivados del mal uso de las tecnologías digitales 170

4.1. El derecho al libre desarrollo de la personalidad y la dignidad 171

4.2. El principio de igualdad y el derecho a la no discriminación . . 174

4.3. El derecho al honor, a la intimidad y a la propia imagen en relación con el derecho a la protección de datos 176

4.4. Derechos de los consumidores 179

4.5. Derecho a la seguridad ciudadana 182

Conclusiones 184

Bibliografía 186

INTELIGENCIA ARTIFICIAL Y CADENA ALIMENTARIA: MÁS ALLÁ DE LA LOGÍSTICA
MARÍA BURZACO SAMPER 191

1. Introducción 191

1.1. Breve apunte sobre el concepto de cadena alimentaria en la legislación 191

1.2. La importancia estratégica del sector agroalimentario. Un apunte sobre sus vulnerabilidades 193

2. IA y cadena alimentaria: las expectativas generadas 196

2.1. Cuestiones generales: las previsiones contenidas en políticas públicas europeas y nacionales 196

2.1.1. Política Agraria: la innovación dirigida a la sostenibilidad ambiental y territorial. 197

2.1.2. Innovación tecnológica e IA 200

2.2. La IA en el sector primario. Su potencial en el desarrollo rural 202

Página

2.3. La IA en la industria alimentaria. Particular mención a la seguridad alimentaria 204

3. Aplicaciones tecnológicas en la cadena alimentaria: algunos ejemplos 205

3.1. Agricultura y ganadería. 205

3.2. Sector pesquero. 206

3.3. Industria agroalimentaria 208

3.4. Plataformas y redes de participación y cooperación. 209

4. Obstáculos y paradojas del escenario. 211

4.1. Dificultades recurrentes: brecha digital y problemas en torno a los datos 211

4.2. Sector agroalimentario y Medio Ambiente: la tendencia a la relocalización 213

Bibliografía 217

Anexos. 223

DESAFÍOS JURÍDICOS DE LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO DE LA PROTECCIÓN DE LA SALUD
OSCAR MONJE BALMASEDA 227

1. La utilización de la inteligencia artificial en el ámbito de la salud: posibilidades, limitaciones y desafíos éticos y jurídicos 228

2. El planteamiento legislativo en materia de inteligencia artificial y responsabilidad civil en la Unión Europea 236

2.1. La responsabilidad civil en el ámbito sanitario. Responsabilidad objetiva y gestión de riesgos 236

2.2. El posicionamiento inicial de la Unión Europea en materia de responsabilidad civil de los robots y los sistemas de inteligencia artificial. 240

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. 249

Bibliografía 255

Página

PARTE SEGUNDA
INTELIGENCIA ARTIFICIAL Y DERECHO MERCANTIL

Propiedad intelectual
MIGUEL MARTÍNEZ MUÑOZ, ANSELMO MARTÍNEZ CAÑELLAS, BORJA SAINZ DE AJA TIRAPU, ANNA BETTONI 261

LA APLICACIÓN DE SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL FUNCIONAMIENTO DEL CONSEJO DE ADMINISTRACIÓN DE LAS SOCIEDADES DE CAPITAL
MIGUEL MARTÍNEZ MUÑOZ 261

1. Introducción. 261

2. La utilización de la Inteligencia Artificial por el consejo de administración y su implicación en los deberes de los administradores 264

3. La Inteligencia Artificial Autónoma como ciberconsejero. . . 274

3.1. La personalidad electrónica 274

3.2. Implicaciones societarias del ciberconsejero 276

4. Aproximación al fenómeno de las DAO 282

Bibliografía 285

ALGUNAS CUESTIONES SOBRE LA APLICACIÓN DE SISTEMAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN LOS PROCEDIMIENTOS CONCURSALES
ANSELMO MARTÍNEZ CAÑELLAS. 289

1. Introducción. Algunas cuestiones en las que la inteligencia artificial puede ser útil actualmente en los procesos concursales 290

2. La redacción de textos jurídicos en materia concursal 292

3. Tratamiento de la Inteligencia Artificial como colaboradora del administrador concursal. 297

4. Valoración de activos y optimización de la venta de activos, o de unidades productivas 300

Página

5. El asesoramiento automático para determinar la insolvencia actual, inminente o probable del deudor 303
5.1. Técnicas de aprendizaje automático en la predicción de insolvencia 304
5.2. Técnicas de aprendizaje profundo en la predicción de insolvencia 307
5.3. Inconvenientes del uso de estos sistemas de inteligencia artificial predictivos 308
Conclusiones 312
Bibliografía 313

INTELIGENCIA ARTIFICIAL Y PROPIEDAD INTELECTUAL: ANÁLISIS DE LAS PRIMERAS CONTROVERSIAS JUDICIALES ENTRE CREADORES DE CONTENIDOS Y PRODUCTORES DE SOLUCIONES DE IA
BORJA SAINZ DE AJA TIRAPU. 319

1. Introducción. 319
2. Estados Unidos de Norteamérica. 322
2.1. Demandas en el ámbito periodístico. 322
A) The New York Times v. OpenAI & Microsoft. . . . 322
B) The Intercept Media, Raw Story Media & Alternet Media v. OpenAI & Microsoft 332
2.2. Demandas en el ámbito literario 334
2.3. Demandas en el ámbito musical. 335
2.4. Demandas en el ámbito fotográfico 337
3. Resto del mundo. 340
*3.1. Sentencia del Tribunal Regional (*Landgericht*) de Hamburgo de 27 de septiembre de 2024 (310 O 227/231)* 340
3.2. Sentencias del Tribunal de Internet de Guangzhou de 8 de febrero de 2024 y del Tribunal de Internet de Beijing de noviembre de 2023 345
Conclusiones 347

Página

REDEFINIENDO LA AUTORÍA: LA INTELIGENCIA ARTIFICIAL Y SUS RETOS PARA LOS DERECHOS DE AUTOR
ANNA BETTONI 349

1. El Reglamento UE sobre la IA. Modelos «generativos» y «generalistas» 349

2. La normativa europea contra el scraping 352

3. La protección de las obras algorítmicas. La relación entre la idea artística y su extrínsecación material 354

4. La autonomía creativa del sistema algorítmico. IA como pincel o IA como pintor? 356

PARTE TERCERA
INTELIGENCIA ARTIFICIAL, CONTRATACIÓN Y RESPONSABILIDAD

NEGOCIACIÓN ALGORÍTMICA Y VOLUNTAD
FRANCESCO RICCI 361

1. Voluntad y decisiones 361

2. Los sistemas de IA se han vuelto capaces de realizar actos de voluntad? 365

3. Voluntad y contrato algorítmico 366

3.1. La hipótesis del sistema de IA como contratista en un sentido sustancial 367

3.2. Los supuestos del sistema de IA como contratista en sentido formal (representante directo) y como tercer árbitro 369

4. (Continuación) El software como elemento documental de la declaración negocial, del contrato algorítmico y del «biancosegno» 373

5. Smart contracts 375

Página

6. (Continuación) Smart contracts que también son contratos algorítmicos 378

Conclusiones 379

IA Y RESPONSABILIDAD CIVIL. MARCO GENERAL
FABIO ADDIS 383

1. El desafío de la Responsabilidad Civil en la era de la inteligencia artificial 383

2. Características e impactos de la inteligencia artificial en el Derecho civil 386

3. Replantear los principios de la Responsabilidad Civil para la IA 388

4. Enfoques normativos sobre la Responsabilidad Civil para los sistemas de IA 390

5. Perspectivas futuras para el Derecho civil en la era digital . . 391

EL ÁMBITO DE PROTECCIÓN DE LA NUEVA NORMATIVA SOBRE PRODUCTOS DEFECTUOSOS
RICARDO PAZOS CASTRO 395

1. Introducción 395

2. Los sujetos protegidos 398

3. Los daños reparables 400

4. La noción de producto 403

5. La noción de defecto 406

5.1. Las legítimas expectativas de seguridad como referencia 407

5.2. Factores a la hora de determinar el carácter defectuoso de un producto 409

5.3. Algunas reflexiones generales sobre los factores anteriores . . . 414

Conclusiones 416

Página

AI LIABILITY: QUESTIONS AND CHALLENGES IN THE AFTERMATH OF THE AI ACT
FRANCESCO PAOLO PATTI 419

1. Introduction 420
2. Aims and strategy of the European Union 422
2.1. Holistic Approach 422
2.2. Concrete instrument for achieving the aims 423
3. Key aspects of the proposed regime 424
4. Relationship to other European instruments 427
5. An Italian perspective on the AILD 430
6. Critical assessment of the AILD 433
6.1. Scope of Application 433
6.2. Fault Liability vs. Strict Liability 434
6.3. The Regulatory Technique 437
Conclusions 439

Inteligencia Artificial y Propiedad Intelectual: Análisis de las primeras controversias judiciales entre creadores de contenidos y productores de soluciones de IA

Borja Sainz de Aja Tirapu
Abogado. Uría Menéndez[1]

SUMARIO: 1. INTRODUCCIÓN. 2. ESTADOS UNIDOS DE NORTEAMÉRICA. *2.1. Demandas en el ámbito periodístico.* A) The New York Times v. OpenAI & Microsoft. B) The Intercept Media, Raw Story Media & Alternet Media v. OpenAI & Microsoft. *2.2. Demandas en el ámbito literario. 2.3. Demandas en el ámbito musical. 2.4. Demandas en el ámbito fotográfico.* 3. RESTO DEL MUNDO. *3.1. Sentencia del Tribunal Regional (Landgericht) de Hamburgo de 27 de septiembre de 2024 (310 O 227/231). 3.2. Sentencias del Tribunal de Internet de Guangzhou de 8 de febrero de 2024 y del Tribunal de Internet de Beijing de noviembre de 2023.* CONCLUSIONES.

1. INTRODUCCIÓN

Estas líneas son continuación del artículo que publiqué con motivo de la primera edición del Congreso Internacional sobre Derecho e Inteligencia Artificial organizado por la Universidad de Deusto y Uría Menéndez. Al igual que la ponencia que tuve ocasión de impartir entonces, ese primer artículo llevaba por título simplemente «Inteligencia Artificial y Propiedad Intelectual» y perseguía el propósito de facilitar una panorámica introduc-

1. Deseo expresar mi agradecimiento a mi compañera María González Fernández por su eficaz y desinteresada ayuda en la recopilación de la información necesaria para la preparación del presente artículo, así como en su análisis y exposición.

toria de las principales cuestiones jurídicas (conflictivas o no) que se plantean al aproximarnos al *machine learning* desde la disciplina de la propiedad intelectual, entendida en sentido amplio.

Entonces no existían aún conflictos judiciales entre los titulares de derechos y los productores de soluciones de inteligencia artificial por esta causa. Tan sólo podían encontrarse en la prensa algunas referencias a posibles controversias en un estado más o menos incipiente y algunas iniciativas, como el Diálogo de la OMPI sobre Propiedad Intelectual e Inteligencia Artificial, iniciado en 2019, para recabar información y opiniones de los operadores concernidos que contribuyeran a fijar posición de cara a la introducción de modificaciones legislativas a nivel global, de cuya necesidad ya entonces se debatía.

Un año después y a medida que las soluciones de inteligencia artificial («IA») se han ido poniendo a disposición de las empresas y, sobre todo, del gran público para la creación de contenidos (en forma de respuestas de texto o generación de imágenes, pistas de audio, vídeos, etc.), han ido surgiendo debates procesales con la interposición de las primeras demandas por la supuesta infracción de derechos de autor y la presunta realización de prácticas desleales en el marco de la creación y explotación de herramientas de IA como Chat GPT, Diffus (Stable Diffusion), Suno AI o Perplexity.

En tanto se concreta el debate legislativo, estas batallas judiciales, que se encuentran con carácter general en fase alegatoria, arrojarán algo de luz en cuestiones como (i) si es posible utilizar sin autorización obras protegidas para el entrenamiento de modelos de IA; (ii) si, independientemente de ello, el funcionamiento y los resultados de las herramientas de IA implican reproducciones de obras sujetas a derechos de propiedad intelectual; (iii) si procede reconocer y en qué medida protección jurídica sobre los resultados generados por la IA; y (iv) quién debe ser, en su caso, el titular de esta protección y el responsable por los posibles incumplimientos.

Precisamente por ello, este artículo se orienta a ofrecer una visión de las posiciones de demandantes y demandados en estas contiendas iniciales así como un adelanto de lo que se ha fallado en las primeras, muy incipientes, resoluciones judiciales. Es aún pronto para vislumbrar el escenario que resultará al final de todos estos procesos. Sin embargo, la exposición de las alegaciones de los creadores de contenidos, por un lado, y de los productores de soluciones de IA, por otro, permitirá trazar cuando menos las lindes del debate jurídico.

La mayoría de estos casos se han originado en Estados Unidos[2], no sólo porque los principales productores de estas herramientas son oriundos de esa jurisdicción sino porque, como cabía esperar, los principales fabricantes de contenidos y, sin duda, las editoriales de prensa cuyas hemerotecas son utilizadas de manera relevante para el entrenamiento de los modelos de IA (como The New York Times, The New York Daily News, etc.) son estadounidenses. Y también porque buena parte de las actividades de entrenamiento, qué duda cabe, se han llevado en ese país. Sea como fuere, en los últimos meses, el debate procesal se ha ido expandiendo a otras jurisdicciones, como cabía esperar. Recientemente, un tribunal alemán ha dictado una sentencia sin precedentes en el ámbito europeo que enmarca la preparación de los datos para el entrenamiento de herramientas de inteligencia artificial dentro de la llamada excepción de minería de datos[3] y, en China, son ya varios los pronunciamientos judiciales que condenan a las empresas desarrolladoras de IA[4].

Este debate global afecta a todos los ámbitos creativos y, principalmente, al periodístico, literario, fotográfico, musical y audiovisual y plantea la necesidad de encontrar el equilibrio adecuado entre dos fuerzas motrices de la sociedad: la cultura y la innovación.

A continuación se expone una muestras de las controversias judiciales más interesantes en la materia.

---

2. La Universidad George Washington ofrece una base de datos con los principales casos en Estados Unidos relacionados con la Inteligencia Artificial - *vid*. https://blogs.gwu.edu/law-eti/ai-litigation-database/, consultada por última vez el 24 de noviembre de 2024. Entre ellos, se encuentran los siguientes:
   — Andersen v. Stability AI Ltd.
   — Authors Guild v. OpenAI Inc.
   — Allen v. OpenAI.
   — GEMA v. OpenAI Ireland Ltd.
   — Chabon et al. V. OpenAI.
   — Daily News, LP v. Microsoft Corp.
   — The Intercept Media, Inc. v. OpenAI, Inc.
   — CanLII v. Caseway AI Legal Ltd.
   — Dow Jones & Company, Inc. v. Perplexity AI, Inc.
   — Toronto Star Newspapers Ltd. v. Open AI Inc.
   — Williams-Sonoma, Inc. v. Amazon.com, Inc.
   — Bartz v. Anthropic PBC.
3. Sentencia del Tribunal Regional (*Landgericht*) de Hamburgo de 27 de septiembre de 2024 (310 O 227/231).
4. Sentencias del Tribunal de Internet de Guangzhou de 8 de febrero de 2024 y del Tribunal de Internet de Beijing de noviembre de 2023.

2. ESTADOS UNIDOS DE NORTEAMÉRICA

2.1. DEMANDAS EN EL ÁMBITO PERIODÍSTICO

La demanda presentada en diciembre de 2023 por *The New York Times* («NYT») contra OpenAI y Microsoft ha inaugurado una serie de reclamaciones de los medios contra estas tecnológicas que dicen tener el objetivo común de preservar la sostenibilidad del periodismo de calidad por medio del respeto a la propiedad intelectual de sus obras.

Estos medios aseveran no estar en contra de las herramientas de IA en sí mismas; persiguen que las empresas desarrolladoras como OpenAI y Microsoft no utilicen sus contenidos sin recabar antes la oportuna autorización de uso de los titulares de los derechos a cambio de una remuneración adecuada que les retribuya por sus esfuerzos en la creación y verificación de los contenidos que publican y, en fin, se busca que la utilización de esos contenidos se lleve a cabo reconociendo su autoría.

Por su parte, las empresas tecnológicas se defienden de esas demandas alegando, entre otros razonamientos, que la utilización de contenidos protegidos para la creación de herramientas de inteligencia artificial generativa sería lícita conforme a la doctrina estadounidense del *fair use*.

A continuación, se exponen con algo más de detalle los casos más representativos y las posiciones de demandantes y demandados.

A) The New York Times v. OpenAI & Microsoft

Posición de NYT

Como adelantábamos, el 27 de diciembre de 2023, el diario estadounidense NYT interpuso demanda contra OpenAI y Microsoft por el supuesto uso no autorizado de miles de artículos protegidos en el desarrollo y utilización de las herramientas de IA ChatGPT y Bing Chat (o «*Copilot*»), respectivamente. Este caso reviste especial importancia por tratarse de la primera demanda presentada por un gran medio de comunicación estadounidense contra empresas generadoras de IA.

Las causas de pedir que ejercita NYT contra los demandados son cinco: (i) infracción directa de derechos de propiedad intelectual por almacenar y reproducir obras protegidas del demandante para el entrenamiento de ChatGPT («infracción de derechos de autor»); (ii) infracción indirecta o

contributory copyright infringement[5] por beneficiarse económicamente de la actividad infractora y, en concreto, por proporcionar Microsoft a OpenAI la infraestructura y recursos para operar y comercializar los modelos de lenguaje («infracción indirecta»); (iii) infracción del artículo 1.202 del *Digital Copyright Millennium Act* («DMCA») por distribuir las obras derivadas sin la información relativa a su autoría y a las advertencias de protección de cada obra («CMI[6]») («infracción del DMCA»); (iv) la infracción de la normativa de competencia desleal por *free riding* o utilización indebida de las prestaciones de NYT para llevar a cabo servicios competidores («infracción de competencia desleal»); y (v) la dilución de la reputación asociada a la marca en la medida en que, en determinados *outputs* de ChatGPT, se atribuyen erróneamente a NYT contenidos de menor calidad o de calidad no verificada por NYT («infracción marcaria»).

En particular, la demanda afirma que, desde 2019, las empresas demandadas habrían venido colaborando estrechamente en la creación de modelos de lenguaje de gran tamaño o *large language models* («LLM»). Según la demandante, esos modelos, en los que se basa ChatGPT, infringirían sus derechos de varias maneras distintas. Por un lado, por haberse creado mediante técnicas de entrenamiento a partir de ingentes cantidades de datos preexistentes que comprenderían contenidos sujetos a los derechos de autor de la demandante sin su autorización. Por otro, porque los *outputs* generados por el chat reproducirían obras del NYT de varias maneras: (i) mostrando copias o reproducciones cuasi-idénticas de las obras originales en respuesta a los *prompts* del usuario; (iii) introduciendo modificaciones

5. Para una síntesis de la doctrina, puede acudirse a la página web de la Copyright Office de Estados Unidos (consultada por última vez el 20 de noviembre de 2024), que hace referencia a varios casos como *Gershwin Publishing Corp. v. Columbia Artists Management, Inc., 443 F.2d 1159, 1162 (2d Cir. 1971) que* la describe en los siguientes términos (énfasis añadido):
«*As the Second Circuit Court of Appeals has explained, contributory infringement occurs where "[o]ne who, with knowledge of the infringing activity, induces, causes, or materially contributes to the infringing conduct of another". In general, the "two elements of contributory infringement" are* (1) "*knowledge*" *of the infringing activity [by a third party]; and* (2) "*material contribution*" *to the activity.*
Note that the common formulation of contributory infringement includes a reference to those who "induce" another to commit infringement, which is similar to the language of the proposed bill. In theory, the existing contributory infringement doctrine may be flexible enough to "cover" situations where a defendant induces another to commit infringement. When actually applying this standard, however, courts have not focused on the inducement aspect of contributory infringement, and thus, [...] "*legislation is needed to address these circumstances*" *directly to remove any doubt about liability for inducement*».

6. CMI es el acrónimo de «*copyright management information*», esto es, información sobre autoría y el contenido de la obra que es difundida junto con ésta para proteger los derechos subyacentes.

no autorizadas en éstas y (ii) dando lugar a resultados de búsqueda que reproducirían extractos de los artículos más allá de los fragmentos (*snippets*) mostrados habitualmente por los buscadores en internet en las respuestas naturales a las búsquedas de los usuarios y que no incluirían hipervínculo a la página oficial de NYT. De lo anterior, NYT extrae además que el modelo en sí debe contener «reproducciones» de las obras (aunque sea en forma de los parámetros que configuran su red neuronal). De otra manera —se afirma— la herramienta no sería capaz de ofrecer resultados que reproduzcan literalmente las obras originales.

Como consecuencia de todo lo anterior, la demandante afirma que, si sus lectores pudieran acceder al contenido publicado por NYT a través de herramientas de IA como ChatGPT o Copilot, no tendrían necesidad de suscribirse a sus servicios informativos. Y ello pondría en peligro su subsistencia. En definitiva, NYT pone sobre la mesa que las demandadas podrían convertirse en los principales y más directos competidores de los medios de comunicación con la ventaja adicional de hacerlo, en buena parte, a costa del esfuerzo de los editores[7].

Nos hallamos aún en un estadio alegatorio del procedimiento por lo que nada ha quedado demostrado todavía. Sin embargo, interesa dejar constancia de los principales hechos sobre los que NYT sustenta su demanda la que hemos llamado «infracción de derechos de autor».

En primer lugar, NYT se refiere a las infracciones que se producirían con motivo de la fase de entrenamiento de ChatGPT. Indica que, según la información pública disponible en relación con ChatGPT-3, los *tokens* (o unidades de información) en que el modelo descompone los contenidos y materiales informativos de los que parte para su entrenamiento tienen las siguientes fuentes:

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Fuente: Cuadro incluido en el apartado 86 de la demanda de NYT en el Case Núm. 1:23-cv-11195 (SHS) (OTW) en relación con las fuentes de ChatGPT-3

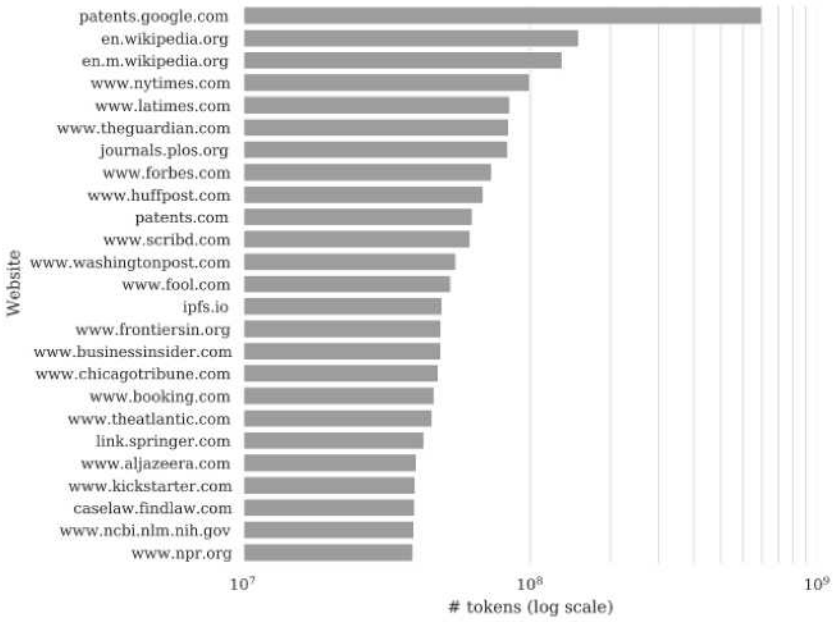
7. https://www.abc.es/tecnologia/the-new-york-times-demanda-openai-microsoft-20231227171512-nt.html# [última consulta realizada el 06 de noviembre de 2024].

Cada una de estas fuentes compone un subconjunto de datos diverso utilizado para el proceso de entrenamiento del modelo de lenguaje ChatGPT-3. Según la actora, la «calidad» de cada subconjunto sería diversa y, correlativamente, su peso ponderado en el «mix» de entrenamiento sería distinta: a mayor calidad, mayor peso ponderado en el algoritmo.

NYT alega que los contenidos de NYT se concentrarían en el *corpus* denominado «WebText2». En particular, afirma que los contenidos de NYT supondrían el 1,23% del total de ese *corpus*. Llama la atención a continuación sobre el hecho de que, pese a que «WebText2» supone menos del 4% del total de *tokens* usados por ChatGPT, tiene un peso ponderado del 22% en la función de entrenamiento. A su juicio, ello es debido a que «WebText2» sería el subconjunto de datos de mayor calidad. Con ello, NYT quiere resaltar el particular valor de sus contenidos para el buen funcionamiento de las herramientas de las demandadas.

NYT apura esta alegación con la descripción que hace a continuación del conjunto principal de datos identificado en la tabla como «Common Crawl». Según NYT, «Common Crawl» sería, en román paladín, una «copia de internet», pero en la que, de nuevo los contenidos de NYT serían relevantes («[t]*he most highly weighted dataset in GPT-3, Common Crawl, is a "copy of the Internet*". [...] *The domain www.nytimes.com is the most highly represented proprietary source (and the third overall behind only Wikipedia and a database of U.S. patent documents) represented in a filtered English-language subset of a 2019 snapshot of Common Crawl, accounting for 100 million tokens (basic units of text)*».[8]).

8. Apartado 88 de la demanda de NYT, disponible en https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.1.0.pdf [última consulta realizada el 26 de noviembre de 2024].



Fuente: Cuadro incluido en el apartado 88 de la demanda de NYT en el Case Núm. 1:23-cv-11195 (SHS) (OTW) en relación con las fuentes de ChatGPT-3

La parte más interesante de la demanda, a los efectos que nos ocupan, es la referente a la descripción del proceso de entrenamiento y funcionamiento[9]. Según NYT, el entrenamiento de modelos de lenguaje como ChatGPT consistiría en «*almacenar copias codificadas de las obras de entrenamiento en la memoria del ordenador para hacerlas pasar repetidamente por el modelo*» a fin de «*ajustar los parámetros para minimizar la diferencia entre las palabras faltantes y las que el modelo predice para rellenar estas*». NYT afirma que los modelos entrenados de este modo mostrarían un comportamiento denominado «memorización» de manera que, ante un *input* o *prompt*, serían capaces de repetir partes importantes de los materiales de entrenamiento. Esto explicaría, según NYT, que ChatGPT sea capaz de reproducir, si así se lo pide el usuario, extensos extractos literales de los artículos de NYT.

Por todo ello, NYT considera que no sólo el proceso de entrenamiento sino también el funcionamiento de ChatGPT conlleva reproducciones no

9. Apartados 78 y ss. de la demanda de NYT, disponible en https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.1.0.pdf [última consulta realizada el 26 de noviembre de 2024].

autorizadas de sus obras. Según NYT, ChatGPT «*codifica copias recuperables de muchos de esos trabajos de entrenamiento*».

Posición de OpenAI

Ni OpenAI ni Microsoft han tenido ocasión procesal de contestar la demanda en cuanto al fondo. Por ahora, sólo han opuesto diversas excepciones procesales (*motions to dismiss*)[10]. En todo caso, tanto por lo que avanzan en éstas como por los documentos que han hecho públicos en otros ámbitos, cabe deducir que su defensa se fundará en la doctrina del *fair use*.

En efecto, el 8 de enero de 2024, *OpenAI* hizo público un comunicado en el que se pronunciaba sobre la demanda interpuesta con un mensaje claro; la conducta objeto de disputa entraría dentro de los límites de la doctrina del *fair use*, ya que el uso de material publicado en internet para entrenar herramientas de IA obedece a un fin «transformador» que sería «*justo para los creadores, necesario para los innovadores y fundamental para la competitividad de Estados Unidos*»[11].

Mayor detalle argumental puede encontrarse en la respuesta de OpenAI a la «Notice of Inquiry and Request for Comment» lanzada a modo de consulta pública por la US Copyright Office en agosto de 2023[12]. A partir de esta respuesta, puede inferirse que OpenAI habría rechazado públicamente determinados hechos en los que se basa la demanda de NYT.

En su respuesta a la US Copyright Office, OpenAI alega en primer lugar que, una vez que se ha concluido el entrenamiento, sus herramientas de IA no guardarían copia de los materiales empleados para tal fin. Con ello, sale

10. En particular, el 26 de febrero de 2024, *OpenAI* solicitó la desestimación de la demanda por diversos motivos preliminares. Con respecto a la infracción de derechos de autor, alegó la prescripción de la acción por haber transcurrido el periodo de tres años para ejercitar acciones civiles, previsto en el artículo 507 del Código de los Estados Unidos o *United States Code* (el «Código»). Con respecto a la infracción indirecta, argumentó que no se cumplirían los requisitos legalmente exigidos por la doctrina del «*contributory infringement*» ya que NYT se refiere en su demanda a un conocimiento generalizado de la infracción por parte de terceros en lugar de a un conocimiento real por *OpenAI* de infracciones específicas. En tercer lugar, defendió que no existe infracción del DMCA porque la demandante no identifica ni cuál es el elemento «CMI» en concreto que había sido eliminado ni cuáles son los hechos que demuestran que *OpenAI* habría actuado con conocimiento de causa (elemento de voluntariedad exigido por la norma). Microsoft solicitó la desestimación de la demanda el 4 de marzo de 2024, por falta de fundamentación en los mismos términos que Open AI.
11. https://www.abc.es/tecnologia/the-new-york-times-demanda-openai-microsoft-20231227171512-nt.html# [última consulta realizada el 06 de noviembre de 2024].
12. Véase https://www.regulations.gov/comment/COLC-2023-0006-8906 [última consulta realizada el 24 de noviembre de 2024].

al paso de la afirmación del NYT de que el modelo sí contendría reproducciones (en forma de parámetros o *tokens*) de obras originales incluso después de haberlas usado para su entrenamiento. Al contrario, OpenAI manifiesta que las reproducciones exactas contenidas en algunas de las respuestas dadas por ChatGPT no obedecen a que éstas se guarden directa o indirectamente por el sistema. OpenAI defiende que cada respuesta es generada *ex novo* cada vez que se le formula una pregunta, sin recuperarla de bases de datos o de otro modo. Si en esas respuestas «novedosas» se contienen reproducciones literales de materiales preexistentes es sólo debido a que determinados textos, por su gran difusión, se encuentran en un mayor número de los materiales de entrenamiento, lo que llevaría al modelo a ser capaz de repetirlos «de memoria» a causa de la relación estadística particularmente intensa existente entre esos datos. De alguna manera, lo que se pone de manifiesto es que, de la misma forma que el modelo detecta que los *tokens* correspondientes a las palabras «gotas» y «lluvia» o bien «mar» y «océano» tienen entre sí una gran proximidad estadística, las palabras de textos recurrentes en la historia o en la cultura guardarían, en un determinado dominio o contexto, una proximidad tal que les haría aparecer en orden literal en el resultado correspondiente, casi como nuestra memoria nos permite repetir versos famosos cuando alguien nos da su entradilla.

Por su interés, reproducimos a continuación los párrafos relevantes de esa respuesta en que OpenAI justifica esta posición:

> «[...] *Developing an advanced model like GPT-4 requires (1) teaching it intelligence, such as the ability to predict, reason, and solve problems, as well as (2) aligning it to human values and preferences. The former is done in a process called "pre-training", which involves showing the model a vast amount of human knowledge using months of supercomputer time. To then incorporate human choice into the model, we use a second step, called "post-training", where we make the model safer and more usable. For a LLM like those powering ChatGPT, pre-training teaches language to the model, by showing the model a wide range of text, and, utilizing sophisticated statistical and computational analysis, having it try to predict the word that comes next in each of a huge range of sequences. This requires an enormous amount of training data and computation, as models review, analyze, and learn from trillions of words* [13].
>
> [...]
>
> *Despite a common and unfortunate misperception of the technology, "the models do not store copies of the information that they learn from". Instead, models are made up of large strings of numbers (called "weights" or "parameters"), which* software *code interprets and executes. The most powerful models consist of billions of weights.*

13. *Ibid*. página 6.

> *Each weight roughly reflects the statistical relationship between different words in different scenarios. As models "learn" during the training process and become better at predicting the next word, their weights update to reflect this improvement. "When asked for a response, the model uses its weights to write a new response each time it is asked". It does not copy its response from its pre-training data, or access it via a database. "Much like a person who has read a book and sets it down", our models do not have access to training information after they have learned from it*[14].
>
> [...]
>
> "*Because our models do not have access to training information after they have learned from it, they are unlikely to duplicate training data in their outputs". In fact, verbatim repetition or "memorization" of training data is generally considered by AI developers to be a bug to be corrected, rather than a feature to be pursued. OpenAI has employed numerous measures to reduce the incidence of this happening, and we regularly update our practices to deploy more. For example, OpenAI takes steps to de-duplicate training data. Showing the model the same information more than once can increase the likelihood of memorization. [...] Nevertheless, memorization can sometimes occur when excerpts of content appear in training data more than once in different contexts and places, such as in reviews, analysis, and commentary. In these cases, an AI model may inadvertently associate certain words or concepts with one another, just as a human might. For example, when someone recites "four score and seven years ago...", it is difficult not to immediately think of "...our fathers brought forth upon this continent, a new nation, conceived in liberty". This is because we have all read and heard these opening words from the Gettysburg Address many times, in many contexts, such that the first six words strongly evoke the remainder of the sentence. OpenAI's language models can exhibit the same behavior where the training data contains the same words appearing in the same order in multiple places and contexts*[15]». [*énfasis añadido*].

En segundo lugar, OpenAI manifiesta que el entrenamiento de ChatGPT no se ordena a reproducir los materiales de partida cuando se lo pide el usuario sino a aprender la relación sintáctica entre sus partes para perfeccionar su modelo predictivo y, en definitiva, para estar en condiciones de generar nuevos resultados, desvinculados de los materiales de partida, ante nuevas preguntas. Por ello, OpenAI afirma que ChatGPT tiene una finalidad transformativa y no reproductiva de la obra original plenamente subsumible en el «uso justo»:

> *«Our models are pre-trained using (1) information that is publicly available on the internet (2) nonpublic information that we obtain from third parties through commercial arrangements, and (3) information that our users or human trainers create and provide*[16].

14. *Ibid*. página 7.
15. *Ibid*. página 8.
16. *Ibid*. página 6.

[...]

"*OpenAI believes that the training of AI models qualifies as a fair use*", *falling squarely in line with established precedents recognizing that the use of copyrighted materials by technology innovators in transformative ways is entirely consistent with copyright law. In considering the fair use question, two interrelated characteristics of model training must be kept in mind. First, "when undergoing pre-training, a model is not interested in the expressive aspects of individual copyrighted works. Instead, as described above, the pre-training process is a highly sophisticated computational process that teaches the model to analyze the structure and syntax of language and images in general terms, to discern the statistical relationships between words, shapes, colors, textures, and concepts". The process used to train an LLM to understand words, concepts, and language is described above, derived at its core from the relational proximity between words that appear in large quantities of text*[17]». [*énfasis añadido*].

En tercer lugar y siempre en relación con el *fair use*, OpenAI niega que su herramienta prive a los autores de la posibilidad de seguir explotando los materiales de partida. Es más, considera que genera, como la mayor parte de las innovaciones tecnológicas, nuevas oportunidades de negocio para los autores y titulares de derechos:

«*Accordingly, while the fair use doctrine remains a keystone of the American AI sector, it will not prevent the formation of a vibrant commercial market for access to, and the display of, copyrighted materials. To the contrary, as described below, OpenAI and numerous content owners with whom we have met believe that generative AI-based products and features "will provide significant new markets for copyrighted works. Particularly exciting in this regard is the potential for enhancing the value of previously overlooked or underutilized collections of copyrighted materials", including archival material. In the specific context of pre-training data, because scale is much more important than popularity or even "quality", previously obsolete and difficult-to-access collections of source code, film archives, amateur video, defunct periodicals, digitized ephemera, and similar materials will likely find a new life in the marketplace.*»[18] [*énfasis añadido*].

Por último y sobre la base de lo anterior, OpenAI argumenta que ChatGPT es una herramienta sustancialmente no infractora, en el sentido de que no se ha concebido para infringir derechos de terceros (sino para resolver necesidades de las personas) y que incorpora numerosos filtros e indicaciones para evitar infracciones. Lo que OpenAI manifiesta no poder evitar es que haya usuarios que fuercen la herramienta para cometer esas infracciones, algo por lo que OpenAI no debería responder salvo de manera secundaria en caso de no reaccionar tras ser advertida de un uso ilícito:

17. *Ibid*. página 12.
18. *Ibid*. página 14.

«In evaluating claims of infringement relating to outputs, the analysis starts with the user. After all, there is no output without a prompt from a user, and the nature of the output is directly influenced by what was asked for. "In OpenAI's experience, the overwhelming majority of users have no interest in infringing any exclusive rights" in any pre-existing copyrighted work. Experience shows that "AI models are not only 'capable of substantial non infringing uses', but are actually used primarily for entirely non-infringing purposes". Moreover, as described above, OpenAI's "models include multiple guardrails" designed to prevent infringing outputs. Nevertheless, a user deliberately seeking to generate an output that includes those elements may, on occasion, be able to evade those guardrails. Here, again, it is the user who determines whether the output implicates the exclusive rights of a copyright owner. In the argot of copyright law, it is the user who is the "volitional actor"[19].

[...]

If infringement has occurred, then copyright law's traditional doctrines of secondary liability can be brought to bear to evaluate any potential liability on the part of the "AI model creator" or service provider. This approach not only reflects the general purpose nature of AI tools and the primacy of the user's contribution, but also accommodates the varied layers of other entities that may be involved. For example, OpenAI makes its models available to other enterprises through its API services. Copyright's secondary liability doctrines are designed to fairly ascribe liability in these contexts, evaluating issues like the accompanying terms of service, control over user conduct, knowledge of infringing outputs, the intent reflected in marketing materials, and the importance of non infringing uses[20]». [*énfasis añadido*].

Recapitulación

Lo expuesto hasta el momento evidencia diferencias radicales entre los proveedores de contenidos y los desarrolladores de soluciones de IA generativa. Para los primeros, no sólo existe un uso no consentido de sus obras como materiales de entrenamiento; el funcionamiento y los *outputs* de esas herramientas también pueden contener reproducciones o transformaciones igualmente infractoras. Y, de ello, deben responder los fabricantes de la herramienta. Para los segundos, las reproducciones técnicamente necesarias para la labor de creación del modelo estarían cubiertas por la doctrina del *fair use* en la medida en que persiguen un avance tecnológico que de otra manera no sería posible y no están diseñadas para reproducir obras preexistentes ni a competir con ellas, sino que, al contrario, generan nuevas oportunidades comerciales. Además, alegan que, después del entrenamiento, la herramienta de IA no conserva reproducciones o transformaciones de los materiales de partida. En fin, consideran que la responsabilidad por las infracciones que se puedan cometer corresponde al usuario que las

19. *Ibid*. página 15.
20. *Ibid*. página 15.

utiliza indebidamente en atención al papel decisivo y no enteramente prevenible de los *prompts* malintencionados. Por ello, los desarrolladores de IA —dicen— sólo deberían responder en la medida en que no adopten medidas de diligencia suficientes para evitar que, una vez conocida, el usuario persista en la infracción.

En suma, este caso plantea algunas de las principales cuestiones inconclusas de la relación entre la inteligencia artificial y la propiedad intelectual. Y, en concreto, las siguientes: ¿Deben ser consideradas las reproducciones de las obras protegidas inocuas a efectos de entrenamiento de soluciones de IA? ¿Suponen los parámetros resultantes de las labores de entrenamiento una forma de reproducción o de transformación de las obras protegidas las soluciones? Y, si lo son, ¿deben quedar igualmente a cubierto en virtud de la doctrina del *fair use*? ¿Hasta qué punto puede una herramienta de IA citar textos o contenidos de terceros en los resultados de sus respuestas? ¿Le es aplicable la doctrina jurisprudencial sobre los *snippets* de los motores de búsqueda? ¿Debe quedar la responsabilidad de los productores de estas herramientas condicionada a su falta de respuesta diligente ante una situación de la que se derive el conocimiento de una situación infractora? Y, de ser así, ¿qué medidas de diligencia debida debería adoptar el proveedor de esas herramientas ante denuncias de uso indebido por usuarios (corrección del modelo, suspensión de la cuenta, etc.)? En fin, ¿Supone una infracción marcaria la errónea atribución del origen empresarial de un determinado contenido?

B) The Intercept Media, Raw Story Media & Alternet Media v. OpenAI & Microsoft

Los medios de comunicación estadounidenses *The Intercept Media*, *Raw Story Media* y *Alternet Media*[21] interpusieron en el 28 de febrero de 2024 demanda contra OpenAI y Microsoft sobre la base de idénticas acusaciones que el NYT. En apretada síntesis, denuncian que las herramientas de estas compañías «*proporcionan o han proporcionado respuestas a los usuarios que reproducen textualmente o casi-textualmente obras periodísticas protegidas por derechos de autor*»[22].

Hay algunas diferencias respecto del caso de NYT.

21. https://storage.courtlistener.com/recap/gov.uscourts.nysd.616536/gov.uscourts.nysd.616536.1.0_1.pdf [última consulta realizada el 21 de noviembre de 2024].
22. https://storage.courtlistener.com/recap/gov.uscourts.nysd.616536/gov.uscourts.nysd.616536.1.0_1.pdf (p. 6) [última consulta realizada el 10 de noviembre de 2024].

En primer lugar, para presentar una demanda por infracción de derechos de autor en Estados Unidos parece necesario que las obras supuestamente infringidas estén registradas en la Oficina de Derechos de Autor de Estados Unidos o *U.S. Copyright Office*. Los medios de comunicación escrita pueden, mediante un único pago mensual, registrar todos los artículos publicados, lo que resulta perfectamente viable para grandes medios de comunicación como NYT. Por el contrario, editores digitales como *The Intercept Media*, *Raw Story Media* y *Alternet* deben presentar solicitudes para cada artículo periodístico que quieran proteger en su sitio web, lo que parece conllevar un coste más difícilmente asumible para editoras más modestas. Por tanto, en vez de acudir a los tribunales por la vía de la infracción de derechos de autor, la estrategia judicial empleada en este caso consiste en alegar una infracción del artículo 1202 del Código —introducido por el DMCA— que se refiere a la prohibición de eliminar la CMI de obras protegidas[23].

Así, las demandantes sostienen que las empresas tecnológicas han infringido la normativa al decidir entrenar sus herramientas de IA a partir de obras periodísticas originales a las que intencionadamente habían retirado este aviso legal. Según las demandantes, OpenAI formó a ChatGPT para que no reconociera ni respetara los derechos de autor, no notificara a los usuarios de ChatGPT cuando las respuestas que recibían estaban protegidas por los derechos de autor de periodistas, y no proporcionara atribuciones cuando utilizara las obras de periodistas humanos.

Hasta la fecha, OpenAI y Microsoft se han pronunciado sobre aspectos meramente procesales instando la desestimación de las demandas con base en los mismos dos argumentos referidos al respecto en la demanda del NYT: de un lado, los demandantes no han conseguido acreditar el daño específicamente ocasionado y, de otro, tampoco han conseguido probar el elemento de intencionalidad para considerar que estamos ante una infracción del artículo 1202 del Código (eliminación de CMI).

A la espera de un pronunciamiento judicial al respecto, a las demandas descritas se suma la presentada de forma colectiva en el mes de mayo de 2024, por un grupo de ocho periódicos estadounidenses —entre ellos The New York Daily News y The Chicago Tribune— contra OpenAI y Microsoft sobre la misma base legal que la del NYT[24].

23. Ibid.
24. https://elpais.com/internacional/2024-05-01/ocho-periodicos-de-estados-unidos-demandan-a-openai-y-microsoft-por-usar-sus-contenidos.html [última consulta 10/11/2024].

2.2. DEMANDAS EN EL ÁMBITO LITERARIO

En la esfera literaria, el caso del sindicato de autores estadounidense, Authors Guild, se dirige a la defensa de los derechos de autor en el contexto del desarrollo de herramientas de IA.

El 19 de septiembre de 2023, el sindicato de autores estadounidense, Authors Guild, junto a diecisiete escritores —entre los que se encuentran George R.R. Martin, John Grisham y Jodi Picoult— presentaron una demanda colectiva contra OpenAI con alegaciones similares a las expuestas en los apartados precedentes. Por excepción, en este caso, no se alega una infracción del DMCA.

Los autores acusan a la compañía de entrenar ChatGPT a costa de una infracción sistemática de sus derechos. En particular, alegan que, hasta poco antes de la demanda, ChatGPT mostraba incluso capítulos completos de obras protegidas por derechos de autor. Por ejemplo, si un usuario quería leer el primer capítulo del libro «*Harry Potter y la piedra filosofal*», de la escritora J.K. Rowling, el *chatbot* reproducía literalmente el primer capítulo de la obra. Ahora, ante la misma consulta, ChatGPT responde de la siguiente manera: «*No puedo mostrarte el texto completo del primer capítulo de Harry Potter y la piedra filosofal, ya que está protegido por derechos de autor. Sin embargo, puedo ofrecerte un resumen*».

Sin embargo, la entidad demandante afirma que el nuevo *output* sigue resultando infractor de los derechos de los autores porque presenta al usuario detalles no accesibles «*ni en reseñas ni en otro material disponible públicamente*»[25], de manera que, a juicio de la actora, el *chatbot* genera en estos casos una obra derivada, fácilmente identificable a partir del original.

La demanda cita el caso de George R.R. Martin, escritor de éxitos literarios como «*Juego de Tronos*». La empresa tecnológica habría utilizado estas obras para entrenar ChatGPT. Como resultado, la herramienta sería capaz no solo de replicar contenido exacto de las novelas sino de redactar de forma minuciosa una secuela alternativa a «*Choque de Reyes*» y titularla «*Un baile con sombras*» o «*A Dance With Shadows*» utilizando los mismos personajes que existen en la serie de novelas de fantasía, «*Canción de hielo y fuego*» o «*A Song of Ice and Fire*». Además, la herramienta también ha sido capaz de crear precuelas y finales alternativos a los ya escritos, con los mismos personajes y escenarios[26].

25. https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.69.0_1.pdf [última consulta realizada el 21 de noviembre de 2024].
26. *Ibid.*, p. 33.

Por otro lado, el sindicato demandante afirma que todas estas infracciones están generando importantes pérdidas para los escritores. A modo de contexto, se cita un estudio realizado recientemente por Authors Guild, conforme al que la mediana de remuneración de los escritores a tiempo completo sería de $20.000 al año. Ello permitiría a Authors Guild justificar que cualquier perjuicio proveniente de la IA es necesariamente inmediato y sustancial[27]. Por otro lado, Authors Guild invoca una encuesta conforme a la que el 69% de los autores preguntados consideran que la IA generativa es una amenaza para su profesión[28].

Tampoco en este caso han tenido ocasión las empresas tecnológicas de defenderse en cuento al fondo del asunto, por lo que, más allá de lo expuesto en los documentos públicos como los citados en los apartados precedentes, desconocemos sus argumentos materiales.

2.3. DEMANDAS EN EL ÁMBITO MUSICAL

En las últimas décadas, la industria musical ha afrontado importantes cambios como consecuencia de la irrupción de plataformas de *streaming* como Spotify o Youtube. La irrupción de la IA plantea ahora nuevos desafíos por su capacidad para componer melodías, escribir letras e incluso imitar el estilo y canciones de artistas conocidos.

UMG Recordings, Sony Music & Warner Music v. Suno Inc. & John Does

Uno de los casos que mejor refleja estos retos es la demanda colectiva interpuesta ante los tribunales estadounidenses por las principales discográficas internacionales, UMG Recordings, Sony Music y Warner Music, contra dos startups de IA, Suno Inc. y John Does.

La herramienta de IA desarrollada por Suno es capaz de generar archivos musicales personalizados en función de los *prompts* de usuarios. De esta manera, cualquier internauta puede convertirse en artista musical en cuestión de segundos simplemente describiendo a la herramienta las características concretas de la canción que se quiere crear.

Sin embargo, la parte actora alega que los resultados de Suno (i) son generados gracias a la previa reproducción no autorizada, en la fase de entrenamiento, de millones de piezas musicales protegidas y (ii) presentan muchas similitudes con grabaciones sonoras preexistentes. En concreto, la

27. Apartado 113 de la demanda disponible en https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.69.0_1.pdf [última consulta realizada el 21 de noviembre de 2024].
28. *Ibid*., apartado 117.

demanda sostiene que «*Suno ha copiado grabaciones específicas con derechos de autor en sus bases de entrenamiento para construir su servicio*»[29].

Esta demanda se refiere expresamente a la doctrina del *fair use* para defender que no resulta aplicable.

La parte actora argumenta que Suno no puede recurrir a la doctrina del *fair use* para justificar lo que considera una apropiación no autorizada y consciente de grabaciones con fines comerciales, especialmente cuando ello conlleva, a su juicio, eliminar la autoría humana, que es la base de la legislación de derechos de autor. Más en concreto, la demanda analiza individualmente los factores de que depende la aplicación de la doctrina del *fair use* y justifica por qué no son de aplicación al caso.

En relación con el primer factor, relativo al propósito y carácter del uso, la actora considera que el realizado por Suno es de naturaleza comercial y además no es transformativo, ya que su modelo de IA crea archivos de música digital que compiten directamente con las grabaciones originales. No sólo ello sino que, a juicio de la demandante, las pistas de sonido generadas por la IA reproducirían con práctica identidad, en muchos casos, las obras preexistentes.

En cuanto al segundo factor (naturaleza de la obra protegida), la demanda sostiene que las grabaciones son expresiones creativas que están en el núcleo de la protección de derechos de autor, lo que dificulta considerar su uso para estos efectos como «justo».

A continuación, y por lo que hace al tercer factor (cantidad y sustancialidad del uso), la actora manifiesta que Suno copia partes importantes y reconocibles de las grabaciones protegidas, lo que inclinaría la balanza, también en este caso, en contra del recurso al *fair use*.

Por último, en cuanto al cuarto factor (efecto en el mercado), la demanda sostiene que el uso de Suno amenaza significativamente el mercado de grabaciones originales y su valor, ya que su servicio podría desplazar o sustituir enteramente las grabaciones originales y a sus autores.

El demandado ha admitido, por su parte, que la herramienta de IA fue entrenada con millones de grabaciones, incluidas algunas de los demandantes. Sin embargo, también en este caso la estrategia defensiva pasa por justificar que su conducta está cubierta por la doctrina del *fair use*.

29. https://www.riaa.com/wp-content/uploads/2024/06/Suno-complaint-file-stamped20.pdf (p. 16) [última consulta realizada el 21 de noviembre de 2024].

En su contestación a la demanda, de 1 de agosto de 2024, Suno argumentó que es legal hacer una copia de una obra protegida como parte de un proceso tecnológico interno que no es visible para el público (el entrenamiento de la aplicación) si persigue la creación de un producto nuevo sustancialmente no infractor. La contestación menciona que hacer una copia intermedia para generar productos no infractores es posible, citando al efecto como ejemplos índices de búsqueda de texto completo y herramientas de detección de plagio, incluso —afirma— a pesar de que el producto final compita directamente con la obra original.

Por otro lado, continúa Suno, bajo la legislación de EE.UU., los derechos exclusivos del propietario de una grabación no se extienden a la creación de otra que consista en una fijación independiente de sonidos incluso si imitan los de la grabación original. En este sentido, el demandado acusa a los demandantes de intentar evitar la competencia utilizando sus derechos de autor. A su juicio, la industria musical ha tratado de bloquear nuevas tecnologías y competidores, y en este caso, los demandantes estarían tratando de impedir el uso de herramientas de IA.

2.4. DEMANDAS EN EL ÁMBITO FOTOGRÁFICO

En Estados Unidos, tras una primera acción a comienzos de enero de 2023 ante los tribunales de Reino Unido contra Stability AI, Ltd. («*Stability UK*»), el 3 de febrero de 2023, la agencia de fotografía neoyorkina Getty Images interpuso demanda contra la empresa estadounidenses Stability AI, Inc. («*Stability AI*») por una presunta infracción de derechos de propiedad intelectual. Ambas demandas, se centran en el uso no autorizado de doce millones de fotografías para entrenar el modelo conocido como Stable Diffusion, una herramienta de IA capaz de generar imágenes por ordenador a partir de una descripción de texto[30].

Getty Images es uno de los principales creadores de contenidos visuales del mundo. Ofrece contenido visual de alta calidad que ayuda a las empresas a reforzar sus marcas. Getty Images almacena sus contenidos en una base de datos que también contiene textos y títulos originales que acompañan a las imágenes individuales (metadatos y descripciones). Así, la galería de fotografías de Getty Images se construye a partir de un conglomerado de creaciones artísticas originales protegidas por la normativa de propiedad intelectual estadounidense, debiendo terceros pagar licencias de uso por cada imagen que quieran descargar.

30. https://storage.courtlistener.com/recap/gov.uscourts.ded.81407/gov.uscourts.ded.81407.1.0.pdf [última consulta realizada el 21 de noviembre de 2024].

Desde su creación en 2020, Stability AI ha centrado su práctica en el desarrollo de modelos de IA para crear contenido digital en línea. Como cualquier otra herramienta de IA, Stable Diffusion necesita nutrirse de ingentes cantidades de datos (*input*) para poder dar respuesta a los *prompts* de los usuarios (*output*). La demanda sostiene que Stability AI copió millones de fotografías —junto con el título y texto asociados a cada una— para entrenar la herramienta, creando un negocio similar al de Getty Images: ofrecer al usuario un resultado visual como respuesta a una descripción de texto.

En cuanto a la CMI, la demanda sostiene que Stable Diffusion habría eliminado intencionadamente esa información de cada una de las imágenes utilizadas para entrenar la herramienta y que la ha modificado intencionadamente proporcionando información falsa al usuario. A partir de pruebas adjuntas a la demanda, Getty Images pretende demostrar que la herramienta de IA ofrecería outputs que, en ocasiones, incluye una versión modificada de la marca de agua propiedad de Getty Images. Sostiene la demanda que esta conducta «*se ha llevado a cabo a sabiendas y con la intención de inducir, permitir, facilitar u ocultar la infracción de derechos de autor*»[31].

Si bien la parte demandada aún no ha tenido oportunidad de pronunciarse sobre el fondo del asunto, Getty Images contrarresta una posible defensa basada en la excepción del *fair use*, alegando en su demanda, tanto el carácter de competidor directo como los cuantiosos beneficios obtenidos por Stability AI, derivados de la actividad ilícita[32].

Mención particular merece la orden judicial dictada el 12 de agosto de 2024 por el tribunal del Distrito Norte de California (*Northen Dictrict of California*)[33] en el marco de una demanda colectiva interpuesta por varios artistas estadounidenses[34] contra las empresas Stability AI, Midjourney Inc. y Deviantart, Inc. y sus herramientas de IA —Stable Diffusion, Midjourney Product y DreamUp, respectivamente— por circunstancias análogas a las ya descritas. En este caso, las empresas demandadas solicitaron la desestimación de la demanda por falta de fundamentación, de acreditación de las

31. https://storage.courtlistener.com/recap/gov.uscourts.ded.81407/gov.uscourts.ded.81407.1.0.pdf (p. 25) [última consulta realizada el 21 de noviembre de 2024].
32. https://institutoautor.org/ee-uu-un-banco-de-imagenes-interpone-una-demanda-contra-un-desarrollador-de-sistemas-de-inteligencia-artificial/ [última consulta realizada el 21 de noviembre de 2024].
33. https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.223.0_2.pdf [última consulta realizada el 26 de noviembre de 2024].
34. Sarah Andersen, Kelly McKernan, Karla Ortiz, Hawke Southworth, Grzegorz Rutkowski, Gregory Manchess, Gerald Brom, Jingna Zhang, Julia Kaye, and Adam Ellis.

concretas obras infringidas (pues la demanda califica erróneamente el *output* generado como obras derivadas) y de identificación del concreto CMI presuntamente eliminado[35].

En concreto, Stability AI alegó la aplicación del *fair use* por crear la herramienta obras completamente originales, alegando que «*Stable Diffusion tampoco es una "herramienta de collage*" [sino que] *permite a los usuarios crear imágenes totalmente nuevas y únicas utilizando sencillas instrucciones de palabras*». Adicionalmente sostiene que no se han identificado las concretas obras que se alegan infringidas[36].

En relación con esas alegaciones, el Tribunal denegó las excepciones orientadas a que la demanda no fuera admitida en relación con la infracción de derechos de autor. El Juez consideró probado a efectos preliminares que el modelo fue entrenado en gran medida por obras protegidas por derechos de autor. Parece que, para ello, el Tribunal se basó principalmente en una declaración del presidente de Stability AI en la que reconocía que el modelo de IA «*tomó 100.000 gigabytes de imágenes y lo comprimió en un archivo de dos gigabytes que puede "recrear" cualquiera de esas imágenes*»[37]. Por tanto, «*Las inferencias plausibles en esta coyuntura son que Stable Diffusion, al ser utilizado por los usuarios finales, infringe los derechos de autor y fue creado para facilitar dicha infracción*»[38].

Además, respecto del argumento sobre la falta de identificación de la titularidad de las concretas obras afectadas, el Tribunal concluye que, dado el gran tamaño de los *datasets* empleados para el entrenamiento de la herramienta, no es necesario que los demandantes presentasen un listado detallado de los contenidos específicamente infringidos[39].

Sin embargo, el Tribunal desestima la alegación por infracción del CMI dado que no se ha probado que fuera eliminado o alterado en relación con copias idénticas a las obras originales, tal y como exige la ley aplicable. Por último, el Tribunal también desestimó las alegaciones por enriquecimiento

35. https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.223.0_2.pdf [última consulta realizada el 26 de noviembre de 2024].
36. *Ibid*., p. 1 [última consulta realizada el 21 de noviembre de 2024].
37. https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.223.0_2.pdf (p. 8) [última consulta realizada el 21 de noviembre de 2024].
38. https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.223.0_2.pdf (p. 9) [última consulta realizada el 21 de noviembre de 2024].
39. *Ibid*.

injusto sobre la base de que estas quedan excluidas del ámbito de aplicación de la normativa de propiedad intelectual porque los demandantes no consiguieron identificar el «elemento adicional» requerido en virtud de la normativa californiana de competencia desleal para cautelar intereses distintos a los protegidos bajo la normativa de propiedad intelectual.

3. RESTO DEL MUNDO

El conflicto sobre el uso no autorizado de fotografías y obras fotográficas protegidas ha traspasado la frontera estadounidense, alcanzando jurisdicciones como la alemana o la china.

3.1. SENTENCIA DEL TRIBUNAL REGIONAL (*LANDGERICHT*) DE HAMBURGO DE 27 DE SEPTIEMBRE DE 2024 (310 O 227/231)

Hasta ahora nos hemos referido a acciones judiciales en Estados Unidos. Era cuestión de tiempo que el debate cruzase el Atlántico, llegando a Europa, donde el conflicto, por ahora, se ha planteado en torno a si el entrenamiento de herramientas de IA puede acogerse al límite de minería de datos introducido por la Directiva (UE) núm. 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre Derechos de autor y derechos afines en el mercado único digital («Directiva DSM»). Como es sabido, esta Directiva regula la minería de datos como límite a los derechos de propiedad intelectual sobre las obras para permitir la generación de conocimiento nuevo sobre patrones, tendencias y correlaciones mediante el análisis a gran escala de los datos accesibles en internet.

En este contexto, el Tribunal Regional de Hamburgo ha desestimado recientemente la demanda interpuesta por el fotógrafo Robert Kneschke contra la entidad alemana sin ánimo de lucro, Laion, por la descarga y uso indebido de sus fotografías para la creación del *dataset* Laion-5B (que ha sido utilizado para entrenar sistemas de IA como Stable Diffussion, entre otros).

Ante las acusaciones de Kneschke, Laion presentó en junio una defensa basada en la aplicación de las excepciones contenidas en los artículos 44a[40], 44b y 60d de la Ley de Derechos de Autor Alemana («UrhG»)[41], que

40. «*Se permiten los actos de reproducción temporal que sean transitorios o incidentales y constituyan una parte integrante y esencial de un proceso técnico y cuya única finalidad sea:*
1. *una transmisión en una red entre terceros por un intermediario o*
2. *un uso lícito*» (traducción propia).

41. https://www.economistjurist.es/articulos-juridicos-destacados/el-fallo-de-hamburgo-un-punto-de-inflexion-en-la-mineria-de-datos-para-entrenamiento-de-la-ia/ [última consulta realizada el 21 de noviembre de 2024].

trasponen al ordenamiento jurídico alemán las disposiciones de la Directiva DSM.

La primera de estas excepciones opuestas por Laion (basada en que las reproducciones serían actos temporales de reproducción) fue rápidamente desestimada por el Tribunal al considerar que la descarga de la imagen por la empresa demandada no podía considerarse un acto accesorio sino deliberado y controlado. Los archivos de imágenes fueron descargados de forma selectiva con el único objetivo de ser analizados mediante un software específico. Ello ejemplifica que, lejos de ser un acto meramente incidental o accesorio, la descarga era una elemento fundamental para llevar a cabo la conducta infractora.

El artículo 44b[42] de la UrhG describe en su apartado primero la minería de datos («*TDM*») como «*el análisis automatizado de obras digitales o digitalizadas, individuales o múltiples, con el fin de obtener información, en particular sobre patrones, tendencias y correlaciones*». De manera relevante, el Tribunal concluyó que la actividad de *scraping* y análisis por la herramienta de IA entra dentro de esta definición, pues su objetivo principal no es otro que encontrar la correlación (es decir, la conformidad o disconformidad) entre las imágenes y la descripción de texto introducida por el usuario, en el sentido literal del precepto.

Además, el archivo de imágenes descargado era también «legalmente accesible» en el sentido del apartado segundo del referido precepto. Este requisito se cumple cuando los usuarios pueden acceder libremente a la obra en internet. En este caso, el demandado descargó la imagen con una marca de agua correspondiente a una imagen de previsualización correspondiente a una librería de imágenes de Internet, de lo que el Tribunal deduce que la obra ya estaba disponible en internet.

Sin embargo, el artículo 44b de la UrhG en su apartado tercero reconoce a los titulares de derechos la posibilidad de excluir el uso de sus obras para la minería de texto y datos, derecho que fue ejercitado por el demandante. En concreto, *Bigstockphoto* —página web en que el demandante publicitaba

42. «*(1) La minería de textos y datos es el análisis automatizado de obras digitales o digitalizadas, individuales o múltiples, con el fin de obtener información, en particular sobre patrones, tendencias y correlaciones.*
(2)Se permiten las reproducciones de obras legalmente accesibles para la minería de textos y datos. Las reproducciones deberán suprimirse cuando ya no sean necesarias para la minería de textos y datos.
(3) El uso conforme al apartado 2, frase 1, sólo está permitido si el titular de los derechos no se ha reservado el derecho de uso. La reserva de uso en el caso de obras accesibles en línea sólo será efectiva si se realiza en formato legible por máquina» (traducción propia).

sus obras— incluía en sus términos de servicio una cláusula que prohibía el uso de programas automatizados para la descarga de su contenido. Dicha reserva, se había realizado respetando los requisitos del articulado, esto es, que fuese clara y legible por una máquina, pese a estar formulada en un lenguaje natural[43]. Por tanto, pese a no pronunciarse sobre la validez del concreto derecho de «*opt-out*» ejercido por la demandante, el Tribunal sí consideró que existían indicios suficientes «*de que en 2021* [fecha de reproducción del acto controvertido] *ya se disponía de sistemas —especialmente para el demandado— capaces de registrar automáticamente una reserva de uso formulada en lenguaje natural*[44]».

La excepción finalmente acogida fue la prevista en el artículo 60d[45] de la UrhG. Este precepto permite la minería de datos realizada por organiza-

43. https://www.economistjurist.es/articulos-juridicos-destacados/el-fallo-de-hamburgo-un-punto-de-inflexion-en-la-mineria-de-datos-para-entrenamiento-de-la-ia/ [última consulta realizada el 21 de noviembre de 2024].

44. Sentencia del Tribunal Regional (*Landgericht*) de Hamburgo de 27 de septiembre de 2024 (310 O 227/231) (p. 8).

45. «*(1) Las reproducciones para la extracción de textos y datos (artículo 44b (1) y (2) frase 1) están permitidas con fines de investigación científica de conformidad con las siguientes disposiciones.*
(2) Las organizaciones de investigación están autorizadas a realizar reproducciones. Serán organismos de investigación las universidades, institutos de investigación u otras instituciones que lleven a cabo investigación científica, siempre que:
1. *no persigan fines comerciales*
2. *reinviertan todos sus beneficios en investigación científica, o*
3. *actúen en interés público en el marco de un mandato reconocido por el Estado.*
Los organismos de investigación que colaboren con una empresa privada que tenga una influencia decisiva en el organismo de investigación y un acceso preferente a los resultados de la investigación científica no tienen derecho en virtud de la frase 1.
(3) También están autorizados a realizar reproducciones:
1. *las bibliotecas y museos, siempre que estén abiertos al público, así como los archivos e instituciones de patrimonio cinematográfico o sonoro (instituciones de patrimonio cultural)*
2. *los investigadores individuales, siempre que no persigan fines comerciales.*
(4) Las personas autorizadas con arreglo a los apartados 2 y 3 que persigan fines no comerciales podrán hacer accesibles al público las reproducciones con arreglo al apartado 1 a las siguientes personas:
1. *un grupo de personas claramente definido para su investigación científica conjunta, y*
2. *terceros individuales con el fin de revisar la calidad de la investigación científica.*
Tan pronto como concluya la investigación científica conjunta o la revisión de la calidad de la investigación científica, se pondrá fin a la puesta a disposición del público.
(5) Las personas autorizadas con arreglo a los apartados 2 y 3 número 1 podrán almacenar copias con arreglo al apartado 1 con las debidas precauciones de seguridad contra usos no autorizados mientras sean necesarias para los fines de la investigación científica o para la revisión de la calidad de la investigación científica.
(6) Los titulares de derechos estarán autorizados a tomar las medidas necesarias para evitar que la seguridad y la integridad de sus redes y bases de datos se vean comprometidas por las reproducciones con arreglo al apartado 1» (traducción propia).

ciones de investigación (como universidades o institutos de investigación) siempre y cuando (i) no persigan fines comerciales (ii) reviertan todos sus beneficios en la comunidad científica o; (iii) desempeñen actividades de interés público por mandato reconocido por el estado alemán.

En opinión del Tribunal, Laion satisface estas características. En primer lugar, define la investigación científica como «*la búsqueda metódica y sistemática de nuevos conocimientos*». Este término debe entenderse en un sentido amplio, de manera que abarque aquellas actividades que tengan por objeto tanto la inmediata como la futura adquisición de conocimiento. Así pues, la creación de un conjunto de datos para entrenar herramientas de IA, si bien no permite, en esa etapa tan temprana, la adquisición de un conocimiento inmediato, sí tiene como objetivo fundamental utilizar la base de datos creada para generar conocimiento futuro.

La falta de ánimo de lucro se evidencia, en opinión del Tribunal, porque el conjunto de datos fue ofrecido a la comunidad de internet por la demandada de forma gratuita. Es destacable que el Tribunal considere, a estos efectos, irrelevante que los servicios de la demandada hayan sido utilizados también por entidades con ánimo de lucro para el entrenamiento de sus propios sistemas de IA. A juicio del Tribunal, esta actividad de las entidades con ánimo de lucro sigue constituyendo una labor de investigación (aun realizada por empresas privadas)[46] y no implica ánimo de lucro por parte de Laion, pues ésta nada obtuvo por esa puesta a disposición.

Por tanto, aunque el fallo reconoce que la imagen utilizada por Laion para la creación de su *dataset* es titularidad del demandante, su reproducción queda amparada bajo la excepción de la minería de texto y datos del artículo 60d de la UrhG[47].

Esta decisión[48] no supone jurisprudencia consolidada pero parece abrir una nueva vía para que las empresas tecnológicas sigan avanzando en el desarrollo de herramientas de IA bajo la excepción de minería de datos.

Con todo, el debate está lejos de quedar resuelto. El 29 de agosto de 2024 se hizo público un dictamen encargado por Initiative Urheberrecht (una organización alemana para la defensa de los derechos de autor) sobre el entrenamiento de modelos generativos de IA, titulado «*Los derechos de Autor y el Entrenamiento de Modelos de IA Generativa - Fundamentos Tecnológicos y Jurídicos*». Con el objetivo de desarrollar una investigación multidisciplinar,

46. Sentencia del Tribunal Regional (*Landgericht*) de Hamburgo de 27 de septiembre de 2024 (310 O 227/231) (p. 8).
47. *Ibid.*
48. https://openjur.de/u/2495651.html [última consulta 10/11/2024].

sus autores —el profesor Dr. Tim W. Dornis, de la Universidad de Hannover y el profesor Dr. Sebastian Stober, de la Universidad de Magdeburgo— han desarrollado pruebas a gran escala que recrean los procesos de entrenamiento de la IA.

Los expertos firmantes consideran que la excepción de la minería de textos y datos no debe aplicarse al entrenamiento de la IA generativa. A su juicio, se trata de realidades tecnológicas diversas. Mientras que la minería de datos «*procesa sólo la información semántica*» la inteligencia artificial generativa «*extrae también la información sintáctica*»[49]. Por otro lado, concluyen que ninguna otra excepción o limitación a los derechos de autor de *lege lata* justifica la reproducción masiva de obras protegidas que tiene lugar durante el entrenamiento de las herramientas de IA. Es más, consideran que, incluso si alguna de ellas fuera aplicable, no se satisfaría la regla de los tres pasos porque la herramienta afectaría a la normal explotación de las obras originales[50]. Ello incluye tanto la recopilación, preparación y almacenamiento de datos protegidos como su posterior reproducción cuasi-idéntica como respuesta a los *prompts* de los usuarios[51]. En todo caso, los expertos recomiendan a los legisladores europeos que el marco jurídico europeo facilite los procesos de entrenamiento para evitar perder competitividad frente a jurisdicciones más liberales pero manteniendo el balance entre la protección de la creatividad humana y el interés para la innovación que supone el desarrollo de la IA[52].

Esta opinión no parece estar, sin embargo, en línea con 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, cuyo considerando 105 afirma que «[l]*as técnicas de prospección de textos y datos pueden utilizarse ampliamente en este contexto* [entrenamiento

49. Copyright Law and Generative AI Training - Technological and Legal Foundations (Urheberrecht und Training generativer KI-Modelle - Technologische und juristische Grundlagen) by Tim W. Dornis, Sebastian Stober :: SSRN [última consulta 10/11/2024]. En particular, se afirma «[e]*l entrenamiento de modelos generativos de IA no limita el uso de los datos de entrenamiento a una evaluación pura de la información semántica contenida en las obras. También capta, en particular, la información sintáctica. Esta evaluación exhaustiva conduce a la representación del contenido de los datos de entrenamiento en el espacio vectorial de los modelos y, por tanto, a una reproducción*» (traducción libre).

50. *Ibid*. «*La evaluación exhaustiva de la información sintáctica de las obras protegidas por derechos de autor debe clasificarse como un menoscabo de la "explotación normal" asignada a los titulares de los derechos por la ley de derechos de autor y, por tanto, como una violación incurable de los requisitos de la prueba*».

51. Copyright Law and Generative AI Training - Technological and Legal Foundations (Urheberrecht und Training generativer KI-Modelle - Technologische und juristische Grundlagen) by Tim W. Dornis, Sebastian Stober :: SSRN [última consulta 10/11/2024].

52. *Ibid*.

de la IA][53] *para la recuperación y el análisis de tales contenidos, que pueden estar protegidos por derechos de autor y derechos afines. Todo uso de contenidos protegidos por derechos de autor requiere la autorización del titular de los derechos de que se trate, salvo que se apliquen las excepciones y limitaciones pertinentes en materia de derechos de autor. La Directiva (UE) 2019/790 introdujo excepciones y limitaciones que permiten reproducciones y extracciones de obras y otras prestaciones con fines de prospección de textos y datos en determinadas circunstancias. Con arreglo a estas normas, los titulares de derechos pueden optar por reservarse sus derechos en relación con sus obras u otras prestaciones para evitar la prospección de textos y datos, salvo que su finalidad sea la investigación científica*».

3.2. SENTENCIAS DEL TRIBUNAL DE INTERNET DE GUANGZHOU DE 8 DE FEBRERO DE 2024 Y DEL TRIBUNAL DE INTERNET DE BEIJING DE NOVIEMBRE DE 2023

El conflicto ha alcanzado también el continente asiático, donde dos recientes fallos chinos se han pronunciado sobre la responsabilidad en que pueden incurrir las grandes empresas tecnológicas por infringir la normativa de propiedad intelectual. Estos casos, aunque menos conocidos globalmente, evidencian la creciente preocupación del gigante asiático por el impacto social, económico y político que las herramientas de IA pueden llegar a alcanzar.

El Tribunal de Internet de Guangzhou (China), se pronunció el pasado 8 de febrero de 2024 (caso 2024 粤 0192 初 113 号),sobre la infracción de derechos de propiedad intelectual en la creación de imágenes generadas por la IA. Este conflicto enfrentó de un lado a Shangai Character License Administrative Co. Ltd —licenciataria de Tsuburaya Productions *Co. Ltd*. que es la entidad propietaria de los derechos de propiedad intelectual del dibujo animado japonés «*Ultraman*»— y de otro a una empresa creadora de una herramienta de IA con funcionalidades imagen-texto[54].

En la demanda se solicitó al Tribunal que se pronunciase sobre la existencia de una infracción de derechos de autor y, en caso de confirmarse, sobre la responsabilidad de la entidad demandada[55].

53. Texto entre corchetes añadido.
54. https://institutoautor.org/china-un-tribunal-se-pronuncia-sobre-la-vulneracion-de-los-derechos-de-propiedad-intelectual-por-la-creacion-de-imagenes-generadas-por-inteligencia-artificial/ [última consulta realizada el 21 de noviembre de 2024].
55. https://copyrightblog.kluweriplaw.com/2024/09/05/awaken-the-warrior-of-light-what-lessons-can-we-learn-from-ultraman/?output=pdf#:~:text=The%20court%20established%20that%20Ultraman,access%20to%20the%20original%20images [última consulta realizada el 10 de noviembre de 2024].

El Tribunal responde afirmativamente a la primera cuestión, considerando que Ultraman es una obra original creada y registrada por Tsuburaya ante la Oficina de Derecho de Autor de China («*National Copyright Administration of China*») y posteriormente licenciada a la demandante. El Tribunal se centra en el *output* generado por la herramienta, concluyendo que esta reproducía imágenes «sustancialmente similares» a las de Ultraman[56].

Declarada la infracción, el Tribunal se pronuncia sobre la presunta responsabilidad de los demandados con base en la normativa china de propiedad intelectual aprobada en noviembre de 2023 («*Interim Measures for the Management of Generative Artificial Intelligence Services*»)[57]. Bajo esta normativa, el Tribunal identificó a la demandada como proveedora de los servicios de IA y por tanto, responsable de adoptar medidas de eliminación, rectificación e información previstas. En este caso, si bien la demandada incluyó filtros de palabras clave que impedían a la herramienta de IA generar imágenes sustancialmente idénticas a las de Ultraman, quedó probado que estas medidas no fueron suficientes[58].

Esta sentencia llega un año más tarde del popular pronunciamiento del Tribunal de Internet de Beijing («*Beijing Internet Court*»), que estimó la demanda interpuesta contra el autor de un blog que había utilizado una fotografía generada por el demandante a partir de IA. El Tribunal consideró que, aun habiendo sido creada con IA, la imagen reflejaba una inversión intelectual del demandante y por tanto, condenó al demandado[59].

56. *Ibid.*
57. Debido a las crecientes disputas originadas en los últimos años por el avance tecnológico, el gobierno chino ha adoptado un carácter proactivo en la regulación de la IA, aprobando en los últimos años, varias iniciativas dirigidas a fortalecer la protección de los derechos de propiedad intelectual dentro de sus fronteras. Las medidas para la gestión de los servicios de inteligencia artificial generativa son el resultado de los esfuerzos coordinados de siete ministerios del gobierno central chino y están dirigidas exclusivamente a las herramientas de IA accesibles a la población china. En concreto, amplían la regulación tradicional en materia de propiedad intelectual imponiendo una serie de obligaciones a los «proveedores de servicio de IA generativa», concepto que incluye a todas aquellas personas que prestan servicios de IA a través de interfaces programables (API). Estas obligaciones abarcan aspectos de gobernanza de datos, etiquetado de contenido generado por IA, protocolos de protección de datos y la protección de los derechos de los usuarios por IA.
58. *Ibid.*
59. https://institutoautor.org/china-ee-uu-se-publican-los-ultimos-pronunciamientos-sobre-la-proteccion-por-el-derecho-de-autor-de-imagenes-creadas-con-la-participacion-de-la-inteligencia-artificial/ [última consulta realizada el 10 de noviembre de 2024].

CONCLUSIONES

Lo expuesto en el presente artículo permite alcanzar algunas conclusiones que son necesariamente provisionales a la vista de los desarrollos legales y jurisprudenciales en curso en esta materia y a los que habrá que estar muy atentos.

1. Los sistemas jurídicos de la inmensa mayoría de países, incluidos aquellos que reconocen el *fair use*, no ofrecen, en cuanto a propiedad intelectual se refiere, la seguridad jurídica necesaria para crear el entorno deseable para el desarrollo de la IA generativa. De otro modo, no se habrían producido en tan corto espacio de tiempo tantos procedimientos judiciales en los que se enfrentan posiciones tan radicalmente irreconciliables.

2. Esta situación no es deseable ni para los creadores de contenidos ni para los fabricantes de IA. Los primeros desconocen si tienen o no a ciencia cierta derecho a impedir y, en su caso, a ser remunerados por la utilización de sus obras para labores de entrenamiento. Y menos aún, conocen cómo, cuándo y en cuánto deben ser compensados. Los segundos, correlativamente, no tienen certeza de si su actividad está o no amparada en los límites a los derechos autor y, en su caso, a quién deberían dirigirse y cómo para obtener, en su caso, los derechos necesarios para llevar a cabo sus innovaciones. Tampoco resulta en absoluto claro que existan mecanismos eficientes para obtener esos permisos y/o efectuar los pagos correspondientes; más bien debe concluirse que no existen en atención a la ingente cantidad de datos que es precisa para perfilar adecuadamente un modelo de IA. En fin, unos y otros deben incurrir en considerables costes judiciales para defender sus respectivas posiciones.

3. Como ha ocurrido tantas otras veces en la historia, ni la incertidumbre legal ni los costes asociados están impidiendo un vastísimo desarrollo de herramientas de IA generativas, cada vez más potentes. El proceso innovador ha sido, es y será imparable. La aprobación del 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, una de las innovaciones legislativas más relevantes a nivel mundial, es buena prueba de que el legislador, sobre todo, reconoce la necesidad de la IA y está interesado en promover su desarrollo.

4. El proceso innovador es, sin embargo, más «imparable» en unas jurisdicciones que en otras. Especialmente, en el ámbito de la Unión Europea, la ausencia del límite de *ius usus inocui* o *fair use* y una prácticamente inoperante regulación de la excepción de minería de datos genera un entorno legal poco propicio para el desarrollo de modelos de inteligencia artificial.

ESTUDIOS

La presente obra se propone como una continuación de la que se publicó hace aproximadamente un año sobre Derecho e inteligencia artificial. En aquella primera ocasión, se trató el tema del jurista frente a los desafíos de la era digital. En este momento, el objetivo es avanzar en el conocimiento de la regulación de la inteligencia artificial. Al igual que entonces, esta obra es fruto de la colaboración entre profesionales, expertos en esta materia, y profesores de diferentes universidades, españolas y extranjeras, que abordan con rigor y profundidad esta materia.

Los contenidos han sido estructurados en tres grandes bloques temáticos. El primero tiene como hilo conductor el marco general regulatorio de la inteligencia artificial y su incidencia en algunos sectores relevantes. El segundo, se centra en el Derecho mercantil y propiedad intelectual. Y el tercero, en la contratación y el Derecho de daños. En suma, una obra completa y ambiciosa que pretende reflejar el impacto de la inteligencia artificial en cada una de estas materias.

El precio de esta obra incluye la publicación en formato DÚO sin coste adicional (papel + libro electrónico)

ACCEDE A LA VERSIÓN ELECTRÓNICA SIGUIENDO LAS INDICACIONES DEL INTERIOR DEL LIBRO

ISBN: 978-84-1162-707-8