
This is the **published version** of the bachelor thesis:

Cabana Puig, Marta. *Models d'intel·ligència artificial de propòsit general i dret d'autor. El procés d'entrenament com a acte de reproducció en el dret de la Unió Europea..* Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Dret). 42 pàg.

This version is available at <https://ddd.uab.cat/record/331857>

under the terms of the  license.



TREBALL FINAL DE GRAU

**Models d'intel·ligència artificial de propòsit general i
dret d'autor.**

El procés d'entrenament com a acte de reproducció en el dret de la Unió Europea.

Autora: Marta Cabana Puig

Tutor: Dr. Santiago Robert Guillén

Grau en Dret

Curs 2025-2026

13 de maig de 2026

Resum

Aquest treball analitza si l'entrenament de sistemes d'intel·ligència artificial generativa pot constituir una infracció del dret de reproducció en el marc del dret d'autor de la Unió Europea, així com l'aplicabilitat de les excepcions de reproduccions provisionals previstes a la Directiva 2001/29/CE i de mineria de textos i dades regulades als articles 3 i 4 de la Directiva 2019/790. Igualment, s'estudien les obligacions de transparència introduïdes pel Reglament 2024/1689 d'intel·ligència artificial i la problemàtica jurídica dels *outputs* generats pels sistemes d'IA.

La investigació s'ha desenvolupat mitjançant una metodologia doctrinal i jurisprudencial, basada en l'anàlisi de normativa europea, jurisprudència del Tribunal de Justícia de la Unió Europea, resolucions judicials nacionals i internacionals i doctrina acadèmica especialitzada, complementada amb una aproximació tècnica al funcionament dels models generatius i dels processos d'entrenament.

L'estudi conclou que l'entrenament de models generatius implica, amb caràcter general, actes de reproducció jurídicament rellevants, mentre que l'excepció de mineria de textos i dades de caràcter general constitueix la principal cobertura jurídica dels desenvolupadors d'IA, tot i les limitacions derivades de l'abast de l'excepció, del mecanisme de reserva de drets i de la manca de transparència dels sistemes.

Paraules clau: dret d'autor, intel·ligència artificial generativa, entrenament, dret de reproducció, mineria de textos i dades, transparència.

Abstract

This paper analyses whether the training of generative artificial intelligence systems may constitute an infringement of the reproduction right under European Union copyright law, as well as the applicability of the temporary copies exception provided in Directive 2001/29/EC and the text and data mining exceptions regulated in Articles 3 and 4 of Directive 2019/790. It also examines the transparency obligations introduced by the Artificial Intelligence Act (Regulation 2024/1689) and the legal issues related to outputs generated by AI systems.

The research has been carried out through a doctrinal and case law methodology, based on the analysis of European legislation, case law of the Court of Justice of the European Union, national and international judicial decisions, and specialised academic literature, together with a technical approach to the functioning of generative models and training processes.

The study concludes that the training of generative models generally involves legally relevant acts of reproduction, while the text and data mining exception constitutes the main legal basis for AI developers, despite the limitations arising from the scope of the exception, the opt-out mechanism, and the lack of transparency of AI systems.

Keywords: copyright, artificial intelligence, training, right of reproduction, text and data mining, transparency.

Abreviatures

Art.: article

DDAMUD: directiva UE 2019/790 sobre els drets d'autor i els drets afins en el mercat únic digital

DPI: dret de propietat intel·lectual

DUE: dret de la Unió Europea

IA: intel·ligència artificial

IAG: intel·ligència artificial generativa

LLM: models de llenguatge de gran escala

LPI: dret de propietat intel·lectual

PI: propietat intel·lectual

STJUE: Sentència del Tribunal de Justícia de la Unió Europea

TDM: mineria de textos i dades

TJUE: Tribunal de Justícia de la Unió Europea

UE: Unió Europea

ÍNDIX DE CONTINGUTS

1. Introducció.....	5
1.1. Context, motivació i rellevància del tema.....	5
1.2. Objectius del treball.....	5
1.3. Metodologia.....	5
1.4. Hipòtesi de treball.....	6
2. La propietat intel·lectual en el DUE	7
2.1. Concepte i funcions, amb incidència en el mercat interior	7
2.2. Configuració dels drets rellevants per a l'entrenament.....	8
2.2.1. Dret de reproducció i actes tècnics previs	9
2.2.2. Drets sobre bases de dades.....	9
2.3. Límits i excepcions rellevants per a usos automatitzats	10
2.3.1. Reproduccions provisionals.....	11
3.2.2. Minería de textos i dades	11
3.2.3. La regla dels tres passos	12
3. Intel·ligència artificial generativa i entrenament.....	13
3.1. Delimitació conceptual mínima per a l'anàlisi jurídic	13
3.2. El procés d'entrenament com a antecedent necessari per identificar actes jurídicament rellevants	14
3.3. <i>Outputs</i> generats, tipologia i rellevància per al risc d'infracció.....	17
4. L'entrenament amb obres protegides per dret d'autor en el DUE	18
4.1. L'entrenament de models com a possible acte de reproducció	18
4.1.1. Reproducció.....	18
4.1.2. Aplicabilitat de l'excepció de reproduccions provisionals.....	20
4.2. L'excepció de minería de textos i dades.....	23
4.2.1. L'excepció de TDM per a recerca científica (article 3 DDAMUD)	23
4.2.2. L'excepció general de TDM (article 4 DDAMUD)	26
4.2.3. Reserva de drets: mecanisme d'opt-out, requisits i efectivitat pràctica	27
4.3. Transparència i compliment normatiu en relació amb continguts d'entrenament	29
5. La dimensió probatòria dels <i>outputs</i> : La generació de continguts com a indicatiu de reproducció en la fase d'entrenament	32
5.1. La regurgitació literal o memorització: L' <i>output</i> com a prova irrefutable de la fixació de l'obra al model	32

5.2. La incorporació d'elements creatius recognoscibles com a prova de la reproducció parcial.....	33
5.3. Anàlisi de la replicació estilística com a frontera de la facultat de reproducció..	34
6. Conclusions	36
7. Llistat de referències.....	38
7.1. Llistat de referències	38
7.2. Jurisprudència	40

1. INTRODUCCIÓ

1.1. CONTEXT, MOTIVACIÓ I RELLEVÀNCIA DEL TEMA

L'entrenament de sistemes d'intel·ligència artificial generativa (en endavant, IAG) amb grans quantitats d'obres protegides planteja importants conflictes amb el dret d'autor, especialment en relació amb la utilització massiva de continguts sense autorització dels titulars de drets i l'abast del dret de reproducció. Aquesta problemàtica adquireix especial rellevància arran dels nombrosos litigis iniciats contra empreses desenvolupadores de models generatius, els quals han situat el debat entre intel·ligència artificial i propietat intel·lectual al centre de l'actualitat jurídica. En aquest context, és interessant l'acord al que va arribar Anthropic, desenvolupadora del model Claude, de 1,5 mil milions de dòlars per resoldre una demanda col·lectiva presentada per diversos autors que acusaven l'empresa d'haver utilitzat milions de llibres protegits per drets d'autor per entrenar els seus sistemes d'IA¹.

En aquest context, el treball neix de la necessitat d'analitzar aquesta problemàtica des d'una perspectiva jurídica rigorosa, marcada per la manca de respostes consolidades i per la tensió existent entre la protecció dels drets dels autors i el foment de la innovació tecnològica. A més, es tracta d'un àmbit especialment recent i dinàmic, en què tant la doctrina com la jurisprudència encara es troben en construcció.

1.2. OBJECTIUS DEL TREBALL

L'objectiu principal d'aquest treball és determinar si l'entrenament de sistemes d'IAG amb obres protegides per drets d'autor constitueix un acte de reproducció en el sentit del dret de la Unió Europea (en endavant, DUE) i, en conseqüència, si pot implicar una infracció dels drets de propietat intel·lectual o quedar emparat pels límits i excepcions existents.

A partir d'aquest objectiu general, es desenvolupen diversos objectius específics: (i) analitzar si l'entrenament de models d'IA pot qualificar-se com a acte de reproducció; (ii) examinar l'aplicabilitat dels límits als drets d'autor, especialment la mineria de textos i dades (en endavant, TDM) i el mecanisme de reserva de drets; (iii) estudiar les obligacions de transparència relatives a les dades d'entrenament; (iv) analitzar el risc d'infracció derivat dels *outputs* generats pels sistemes d'IA; i (v) oferir una valoració crítica sobre l'adequació del marc normatiu europeu davant aquests reptes.

1.3. METODOLOGIA

El treball adopta una metodologia jurídica basada principalment en l'anàlisi doctrinal i jurisprudencial. S'han utilitzat articles acadèmics especialitzats, manuals de dret de propietat intel·lectual (en endavant, DPI), informes institucionals i normativa europea rellevant per construir el marc teòric de l'estudi. Així mateix, s'ha analitzat jurisprudència tant del Tribunal de Justícia de la Unió Europea (en endavant, TJUE) com de tribunals

¹ Jamali, L. (6 de setembre de 2025). *AI firm Anthropic agrees to pay authors \$1.5bn for pirating work*. BBC News. <https://www.bbc.com/news/articles/c5y4jpg922qo>

nacionals i internacionals, especialment dels Estats Units i del Regne Unit, amb la finalitat d'incorporar una perspectiva comparada.

Cal destacar que la limitada existència de resolucions judicials fermes respon al caràcter recent de la matèria, ja que la majoria dels litigis actuals encara no han estat resolts. Aquesta circumstància justifica el pes rellevant de la doctrina i també l'extensió del treball. En efecte, la manca d'una interpretació jurídica consolidada obliga a desenvolupar una anàlisi més detallada i matisada, amb l'objectiu d'oferir una visió completa i rigorosa d'una problemàtica encara en construcció.

També s'ha incorporat una aproximació tècnica al funcionament dels models generatius per comprendre millor les problemàtiques jurídiques analitzades, així com l'ús puntual d'eines d'IA com a suport de recerca i objecte d'observació pràctica, sense substituir el raonament jurídic propi de l'autora.

1.4. HIPÒTESI DE TREBALL

La hipòtesi de partida es basa en que l'entrenament de sistemes d'IAG pot constituir, en determinades circumstàncies, un acte de reproducció jurídicament rellevant que no queda plenament emparat pels límits i excepcions del DUE. En conseqüència, el marc normatiu actual presenta certes insuficiències per donar resposta als reptes plantejats per la IA.

A partir d'aquesta hipòtesi, el treball defensa la necessitat d'una interpretació evolutiva del dret d'autor que permeti mantenir un equilibri adequat entre la protecció dels drets dels autors i la promoció de la innovació tecnològica, evitant tant una restricció excessiva del desenvolupament de la IA com una desprotecció dels interessos legítims dels titulars de drets.

2. LA PROPIETAT INTEL·LECTUAL EN EL DUE

2.1. CONCEPTE I FUNCIONS, AMB INCIDÈNCIA EN EL MERCAT INTERIOR

La propietat intel·lectual (en endavant, PI) són el conjunt de drets que protegeixen les creacions de la ment humana i que permeten als seus titulars gaudir dels beneficis derivats de la seva obra (Organización Mundial de la Propiedad Intelectual, 2020). D'acord amb l'article 2 del Conveni d'Estocolm de 1967, la PI es divideix en dues categories: els drets d'autor i els drets de propietat industrial. Els drets d'autor inclouen les obres literàries, cinematogràfiques, musicals, artístiques i els dissenys arquitectònics. No obstant, en l'àmbit del dret europeu i en alguns ordenaments nacionals com l'espanyol, el terme PI s'utilitza en un sentit més estricte per referir-se específicament al dret d'autor i als drets afins (Candelario Macias et al., 2017).

En el marc de la UE, el dret d'autor ha estat objecte d'un procés d'harmonització progressiva amb la finalitat de garantir el funcionament del mercat interior i evitar distorsions derivades de les diferències entre ordenaments nacionals. En aquest sentit, la Directiva 2001/29/CE sobre la societat de la informació (InfoSoc) constitueix el pilar fonamental del sistema europeu de dret d'autor, establint un marc comú per a la protecció de les obres en l'entorn digital i assegurant un alt nivell de protecció dels titulars de drets, alhora que facilita la lliure circulació de continguts dins de la Unió (EUIPO, 2025).

Des d'un punt de vista conceptual, el dret d'autor es fonamenta en la noció d'obra (arts. 10 i 11 LPI). S'entén per obra tota creació original literària, artística o científica, fruit de l'activitat intel·lectual humana (Candelario Macias et al., 2017; Erdozain López et al., 2023). L'obra és immaterial, no s'identifica amb el seu suport material (paper, llenç, arxiu digital, etc.), encara que normalment requereix d'un suport per existir, conservar-se i ser explotada. No es pot desconèixer, tal i com recorda la jurisprudència del TJUE i especialment la sentència recent de 4 de desembre de 2025, assumptes acumulats C-580/23 i C-795/23, que el concepte d'obra constitueix una noció autònoma que ha de ser interpretada de manera uniforme i exigeix dos elements acumulatius: d'una banda, l'existència d'un objecte original que constitueixi una creació intel·lectual pròpia de l'autor i, de l'altra, que aquesta creació es manifesti a través d'elements que l'expressin. Així doncs, l'element central que determina la protecció és el requisit d'originalitat, entès com el reflex de la personalitat de l'autor mitjançant decisions lliures i creatives, mentre que queden excloses aquelles creacions determinades exclusivament per consideracions tècniques, regles o exigències que no deixen marge a la llibertat creativa (Erdozain López et al., 2023; Rogel Vide & Serrano Gómez, 2021). A més, l'art. 13 LPI disposa que no són objecte de PI les lleis, les resolucions judicials i els actes o dictàmens dels organismes públics, així com les seves traduccions oficials.

Pel que fa al subjecte del dret, la titularitat originària correspon a l'autor, entès com la persona física que crea l'obra. Encara que en determinats supòsits la titularitat pot

correspondre a persones jurídiques, com en el cas de les obres col·lectives o dels programes d'ordinador (Rogel Vide & Serrano Gómez, 2021). L'ús d'eines tecnològiques no exclou l'autoria humana quan actuen com a instruments creatius, però en queden fora les produccions generades de manera autònoma per sistemes d'IA sense intervenció creativa humana significativa, tal com ha assenyalat la *US Copyright Office* en el cas *Zarya of the Dawn* (2023) (Díaz Noci, 2023).

Des d'una perspectiva funcional, la PI es caracteritza per una doble dimensió econòmica i cultural. D'una banda, desenvolupa una funció econòmica en la mesura que configura un sistema d'incentius que permet als titulars obtenir un rendiment de l'explotació de les seves creacions, afavorint la inversió en activitats creatives, tecnològiques i industrials, així com el creixement de sectors estratègics dins del mercat interior. D'altra banda, compleix una funció cultural, ja que contribueix a la producció, difusió i preservació de continguts creatius, promou la diversitat cultural i facilita la circulació d'obres dins de la UE, amb una incidència directa en el desenvolupament social i en l'accés al coneixement (EUIPO, 2025; Fernandes, 2024; Vercelli, 2022).

Aquesta doble dimensió és especialment rellevant en l'entorn digital, on les successives innovacions tecnològiques han transformat profundament les formes de creació, distribució i accés als continguts. L'explotació de les obres es produeix de manera transfronterera i requereix un marc jurídic coherent que garanteixi seguretat jurídica (Sag & Yu, 2025). Així, el dret d'autor ha volgut evolucionar per adaptar-se a aquests canvis, mantenint un equilibri entre la protecció dels titulars i la necessitat de promoure la innovació i el desenvolupament tecnològic (EUIPO, 2025).

2.2. CONFIGURACIÓ DELS DRETS RELLEVANTS PER A L'ENTRENAMENT

El DPI es configura com un conjunt de facultats atribuïdes a l'autor amb la finalitat de protegir tant els interessos econòmics derivats de la seva explotació com el vincle personal que uneix el creador amb la seva obra. Aquest sistema respon a una concepció personalista de la creació, segons la qual l'obra és expressió de la personalitat de l'autor (Candelario Macias et al., 2017). Aquesta idea explica la importància central dels drets morals que garanteixen un vincle jurídic i permanent entre l'autor i la seva creació; en contrast amb el model anglosaxó del copyright, tradicionalment orientat a una funció més econòmica i utilitarista (Erdozain López et al., 2023; EUIPO, 2025).

En aquest treball, l'anàlisi se centra en els drets patrimonials exclusius, en concret en el dret de reproducció, que atorguen a l'autor la facultat d'autoritzar o prohibir determinats usos de l'obra i, per tant, d'exercir un control efectiu sobre la seva explotació econòmica (Erdozain López et al., 2023). En aquest sentit, es tracta de drets transmissibles i de durada limitada: s'estenen, com a regla general, durant tota la vida de l'autor i setanta anys després de la seva mort. Un cop extingits, l'obra passa a formar part del domini públic (Candelario Macias et al., 2017).

2.2.1. Dret de reproducció i actes tècnics previs

Entre els drets patrimonials exclusius rellevants per a l'entrenament de sistemes d'IA, el dret de reproducció ocupa una posició central. A nivell internacional, aquets dret es reconeix a l'art. 9 del Conveni de Berna i també es troba harmonitzat en el DUE a l'art. 2 de la Directiva Infosoc (art. 18 LPI), que reconeix als autors i a determinats titulars de drets afins la facultat exclusiva d'autoritzar o prohibir la reproducció de les seves obres, amb independència del caràcter temporal o permanent de la còpia, del mitjà utilitzat i del fet que es tracti de la totalitat o només d'una part de l'obra (Cernadas Lázare et al., 2024; EUIPO, 2025).

Aquesta definició amplia reflecteix la voluntat d'incloure tant les còpies materials tradicionals com les reproduccions digitals, incloses aquelles que es produeixen de manera automàtica en el funcionament dels sistemes tecnològics. Des d'aquesta perspectiva, es consideren actes de reproducció operacions com la digitalització d'una obra analògica o la seva introducció en un servidor, atès que impliquen una fixació de l'obra en un nou suport (Candelario Macias et al., 2017; Erdozáin López et al., 2023).

En aquest context, resulta especialment rellevant la distinció entre actes de reproducció i altres categories jurídiques com la transformació. Així, operacions com la digitalització, la codificació o el fraccionament tècnic de les obres per a la seva transmissió a través de la xarxa no impliquen una aportació creativa nova i, per tant, han de qualificar-se com a reproduccions i no com a transformacions (Candelario Macias et al., 2017; Erdozáin López et al., 2023).

Finalment, cal destacar que moltes de les reproduccions que es produeixen en entorns digitals tenen un caràcter tècnic o instrumental, en la mesura que són necessàries per al funcionament dels sistemes informàtics. Aquest fet és especialment rellevant en el cas de l'entrenament de sistemes d'IA, les diferents fases del procés (des de la recopilació de dades fins al seu tractament i anàlisi) comporten, en major o menor mesura, la captació, emmagatzematge i tractament massiu de dades. Aquestes operacions impliquen la realització de múltiples còpies de les obres o de fragments d'aquestes (EUIPO, 2025; Fernandes, 2024; Li et al., 2024).

2.2.2. Drets sobre bases de dades

A més del dret de reproducció, l'entrenament de sistemes d'IA pot incidir en els drets sobre bases de dades, especialment quan els conjunts de dades utilitzats es configuren com a col·leccions estructurades de continguts protegits (Contardi, 2025). En el DUE, aquests drets es regulen a l'art. 1, Directiva 96/9/CE, que defineix una base de dades com "una col·lecció d'obres, dades o altres materials independents disposats de manera sistemàtica o metòdica i individualment accessibles per mitjans electrònics o d'una altra mena" (EUIPO, 2025). En l'ordenament intern es troben regulats a l'art. 12 LPI.

Aquest règim estableix una doble via de protecció. D'una banda, les bases de dades poden quedar protegides pel dret d'autor quan la selecció o disposició dels seus continguts constitueix una creació intel·lectual pròpia de l'autor. En aquest cas es protegeix l'estructura, i l'autor disposa del dret exclusiu d'autoritzar la reproducció de la

mateixa, tant temporal com permanent (Candelario Macias et al., 2017; Rogel Vide & Serrano Gómez, 2021). D'altra banda, existeix un dret *sui generis* que protegeix la inversió substancial, qualitativa o quantitativa, realitzada en l'obtenció, verificació o presentació dels continguts, però no es tracta en absolut d'una protecció de l'obra com a tal. Aquesta atorga al fabricant de la base de dades la facultat d'impedir l'extracció o reutilització de la totalitat o d'una part substancial dels continguts (Contardi, 2025; EUIPO, 2025; Izquierdo, 2021).

Finalment, el Reglament 2023/2854 sobre accés just a les dades ha introduït una precisió rellevant en aquest àmbit, en establir que les dades generades o obtingudes a través de dispositius connectats a l'Internet de les coses (IoT) no poden beneficiar-se de la protecció del dret *sui generis* sobre bases de dades, limitant així l'abast d'aquesta figura en determinats contextos tecnològics (EUIPO, 2025). És a dir, una empresa no pot utilitzar el dret de bases de dades per impedir que altres accedeixin o reutilitzin dades generades automàticament per dispositius.

En definitiva, en el context de l'entrenament de sistemes d'IA, aquest règim és especialment rellevant, atès que aquests sistemes es basen sovint en grans conjunts de dades estructurades i la seva utilització pot implicar actes d'extracció o reutilització dels continguts protegits (Contardi, 2025). Així doncs, tant el dret de reproducció com els drets sobre bases de dades constitueixen els principals punts de connexió entre el règim de PI i els processos d'entrenament de sistemes d'IA, atesa la naturalesa tècnica i massiva de les operacions que aquests comporten (EUIPO, 2025).

2.3. LÍMITS I EXCEPCIONS RELLEVANTS PER A USOS AUTOMATITZATS

Malgrat l'amplitud del dret de reproducció, aquest no és un dret absolut, el DUE estableix un sistema de límits i excepcions destinat a equilibrar els interessos dels titulars amb els de la societat. Aquestes excepcions permeten determinats usos d'obres protegides sense autorització del titular, per tal d'equilibrar els drets d'autor amb altres interessos generals com la informació, l'educació o la investigació; els límits també permeten l'ús mitjançant un dret a remuneració (Bragado Herrero de Egaña, 2023). Es troben recollits principalment als arts. 31 a 39 i 40 bis LPI (procedents de l'art. 5 de la Directiva 2001/29/CE), han d'interpretar-se de manera restrictiva i conforme a la regla dels tres passos, de manera que en cap cas puguin causar un perjudici injustificat als interessos legítims de l'autor ni afectar l'explotació normal de l'obra (Candelario Macias et al., 2017; Ordellin Font, 2025).

En canvi, a altres sistemes com l'anglosaxó o l'estatunidenc regeix un sistema obert de límits als drets d'autor, conegut com a *fair dealing* i *fair use*, respectivament. Aquest sistema permet considerar lícits aquells usos d'obres protegides que, després d'una valoració cas per cas, superin favorablement una sèrie de criteris. Entre aquests criteris es tenen en compte: la finalitat i el caràcter de l'ús realitzat, la naturalesa de l'obra utilitzada, la proporció i rellevància de la part utilitzada en relació amb el conjunt de l'obra, així com l'impacte que aquest ús pot tenir sobre el mercat potencial i el valor econòmic de l'obra original (Erdozain López et al., 2023; Vercelli, 2022).

2.3.1. Reproduccions provisionals

Una de les limitacions més rellevants del dret de reproducció en l'entorn digital és l'excepció relativa a les reproduccions provisionals, prevista a l'art. 5.1 de la Directiva Infosoc. D'acord amb aquest precepte, queden exemptes del dret de reproducció aquelles còpies que compleixin de manera acumulativa una sèrie de requisits estrictes: han de ser provisionals, transitòries o accessòries, formar part integrant i essencial d'un procés tecnològic, tenir com a única finalitat facilitar una transmissió en xarxa entre tercers o una utilització lícita de l'obra, i no tenir una significació econòmica independent. Es tracta, per tant, d'una excepció funcional, concebuda per permetre operacions tècniques indispensables com la navegació o l'emmagatzematge temporal (memòria cau), sense les quals el funcionament eficient d'internet seria inviable (EUIPO, 2025; Iglesias Posse, 2017).

La jurisprudència del TJUE ha precisat l'abast d'aquesta excepció, i ha considerat lícites les còpies en pantalla i les còpies cau generades durant la consulta de continguts a Internet, sempre que siguin automàtiques, de durada limitada i indispensables per al procés tècnic. En canvi, ha exclòs l'aplicació de l'excepció en aquells supòsits en què les còpies no són realment transitòries, depenen de decisions humanes o es vinculen a usos no autoritzats de les obres (Candelario Macias et al., 2017; Erdozáin López et al., 2023).

3.2.2. Minería de textos i dades

La Directiva 2019/790 sobre els drets d'autor i els drets afins en el mercat únic digital (en endavant, DDAMUD) introdueix excepcions específiques per al TDM, que permeten la reproducció d'obres protegides sense autorització quan tenen per finalitat l'anàlisi automatitzada de textos i dades per extreure patrons, tendències o correlacions (Izquierdo, 2021). Aquest règim reflecteix la voluntat del legislador europeu d'adaptar el dret d'autor a les noves tecnologies, tot mantenint un equilibri entre innovació i protecció (Rosati, 2024).

Concretament, l'art. 3 estableix una excepció obligatòria per a finalitats de recerca científica, mentre que l'art. 4 preveu una excepció més general. Amb caràcter general, no és necessari el consentiment dels titulars sempre que l'accés a les obres sigui lícit. No obstant, les diferències rellevants entre ambdós règims es troben en la durada de la conservació de les còpies i en la possibilitat d'exclusió per part dels titulars de drets. En el cas de l'art. 3, es permet que aquestes es conservin amb un nivell adequat de seguretat i, en particular, per a la verificació dels resultats de la investigació. En canvi, en l'art. 4, les còpies poden mantenir-se durant el temps estrictament necessari per dur a terme el procés de TDM (Serranía, 2020). A més, en aquest segon supòsit, els titulars de drets poden exercir una reserva expressa per excloure l'ús de les seves obres (Erdozáin López et al., 2023; García Vidal, 2024; Izquierdo, 2021).

Aquest règim s'estén també a bases de dades i programes d'ordinador, permetent determinats actes de reproducció, extracció o transformació sense autorització, sempre que es realitzin amb finalitats de TDM (Erdozáin López et al., 2023).

3.2.3. *La regla dels tres passos*

La regla dels tres passos, recollida a l'art. 40 bis LPI, constitueix un criteri clau per a la interpretació i aplicació dels límits als drets d'explotació. Aquesta regla prové tant dels principals tractats internacionals en matèria de dret d'autor com de la normativa europea, en concret es troba a l'art. 5.5 de la Directiva 2001/29/CE (Erdozáin López et al., 2023). La regla estableix que qualsevol límit només és admissible si compleix tres requisits acumulatius: ha d'aplicar-se a casos concrets i especials expressament previstos pel legislador, no pot afectar l'explotació normal de l'obra i no ha de causar un perjudici injustificat als interessos legítims dels titulars.

El primer requisit implica que els límits són tancats i han d'interpretar-se de manera restrictiva, sense que els tribunals puguin crear nous supòsits no previstos per la llei. El segon pretén evitar usos abusius que substitueixin o perjudiquin de manera significativa els ingressos del titular, és a dir, que buidin de contingut els drets d'explotació. El tercer requisit exigeix valorar si el perjudici al titular es justifica per altres interessos, com drets fonamentals o finalitats d'interès general. Finalment, aquesta regla actua com un límit als propis límits i opera tant com a criteri interpretatiu per als jutges com una guia per al legislador en la configuració de les excepcions (Erdozáin López et al., 2023).

3. INTEL·LIGÈNCIA ARTIFICIAL GENERATIVA I ENTRENAMENT

L'estudi del dret de reproducció en el context de la IAG exigeix una anàlisi rigorosa de la fase de configuració dels models, per tal d'identificar quins actes poden ser jurídicament rellevants des del dret d'autor. La IAG no és només una eina de processament d'informació, sinó un sistema que basa la seva funcionalitat en la ingestió massiva d'obres protegides (EUIPO, 2025).

3.1. DELIMITACIÓ CONCEPTUAL MÍNIMA PER A L'ANÀLISI JURÍDIC

L'article 3 del Reglament 2024/1689 d'IA defineix la IA com:

Un sistema basat en una màquina dissenyat per funcionar amb diferents nivells d'autonomia i que pot mostrar capacitat d'adaptació després del seu desplegament, i que, per a objectius explícits o implícits, infereix a partir de la informació d'entrada que rep la manera de generar resultats de sortida, com ara prediccions, continguts, recomanacions o decisions, que poden influir en entorns físics o virtuals.

La IAG és una subcategoria d'aquests models orientada a la creació autònoma de continguts nous, únics i originals, com ara textos, imatges, sons, vídeos o codis informàtics, a partir d'un procés previ d'anàlisi massiva de dades. La IAG no es limita a operar sobre dades existents, sinó que es caracteritza per la generació de resultats diferents a partir dels patrons apresos durant l'entrenament. L'objectiu de l'entrenament de la IAG és desenvolupar models capaços de produir resultats similars a les dades d'entrada (Dornis, 2025). Des d'un punt de vista tècnic, aquests sistemes es basen principalment en tècniques d'aprenentatge automàtic (*machine learning*) i aprenentatge profund (*deep learning*), que utilitza xarxes neuronals artificials per processar grans volums d'informació (Montaño, 2022; Vela, 2024).

Les xarxes neuronals són estructures matemàtiques organitzades en múltiples capes successives que processen la informació de manera jeràrquica, permetent captar patrons simples en les primeres capes i patrons progressivament més complexos en les capes posteriors (*Getty Images v. Stability AI* (2025)). El mecanisme es pot descriure com un procés d'inferència probabilística: davant una instrucció de l'usuari (*prompt*), el model calcula quina és la seqüència de sortida més probable d'acord amb els patrons apresos durant l'entrenament (*Kadrey v. Meta Platforms* (2025)).

Un exemple paradigmàtic d'IAG són els models de llenguatge de gran escala (LLM). Aquests models estan dissenyats específicament per comprendre i generar text en llenguatge natural, i el seu entrenament es realitza sobre enormes col·leccions de textos digitals (EUIPO, 2025; Fernandes, 2024). Entre els models de llenguatge de gran escala més coneguts es troben ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic) o Copilot (GitHub/OpenAI), així com models generatius d'imatge com Stable Diffusion o Midjourney (Sag & Yu, 2025; Vela, 2024).

A nivell normatiu, el marc europeu aplicable és dual. D'una banda, la DDAMUD té com a objectiu modernitzar el dret d'autor a la UE, afavorir l'accés transfronterer als continguts digitals i equilibrar els interessos dels diferents actors en l'economia digital. Dins d'aquest

marc, introdueix dues excepcions específiques per a la TDM, recollides als arts. 3 i 4 (Montaño, 2022; Sag & Yu, 2025).

D'altra banda, el Reglament d'IA parteix de la idea que aquests models incideixen directament en els interessos dels titulars de drets, atès que alteren les formes de producció i ús dels continguts creatius. Per aquest motiu, estableix com a principi general que l'explotació d'obres protegides exigeix el consentiment del titular. Tanmateix, admet que aquest ús pot considerar-se lícit quan es fonamenti en les excepcions del dret d'autor del DUE, especialment la de TDM prevista a la DDAMUD, sempre que es compleixin les condicions i límits corresponents (Ordell Font, 2025). En aquets sentit, introdueix obligacions específiques per als proveïdors de models d'IA de propòsit general i exigències de transparència per garantir el respecte del dret d'autor en els sistemes d'IA (EUIPO, 2025; Navas Navarro, 2025).

L'informe de l'EUIPO (2025) destaca que, des de la perspectiva jurídica, cal distingir dues fases fonamentals: l'ús de continguts com a *input* durant l'entrenament i la generació de continguts com a *output*.

3.2. EL PROCÉS D'ENTRENAMENT COM A ANTECEDENT NECESSARI PER IDENTIFICAR ACTES JURÍDICAMENT RELLEVANTS

L'entrenament d'un model d'IAG és una successió de fases on cadascuna pot constituir un acte de reproducció independent segons l'art. 2 de la Directiva InfoSoc. L'*input*, les dades d'entrada que es proporcionen a un sistema durant la fase d'entrenament, poden tenir diferents formes: seqüències de text, conjunts d'imatges, dades numèriques; depenent de la naturalesa de la IA que s'està entrenant (Cernadas Lázare et al., 2024). L'EUIPO (2025) descriu aquest procés a partir de diverses fases: recollida de dades, accés i ingestió, pre-processament, representació interna i entrenament pròpiament dit. Cadascuna d'aquestes etapes pot tenir rellevància jurídica pròpia, especialment quan implica reproduccions d'obres o extraccions de contingut protegit.

1r. Recollida de dades

En la fase inicial de recollida de dades, la major part d'aquestes dades s'extreuen mitjançant tècniques d'extracció d'Internet (*web scraping*), s'utilitzen rastrejadors per a la descàrrega automatitzada i sistemàtica d'obres i continguts disponibles en línia per tal d'emmagatzemar-los en servidors on podran ser analitzats, depurats i filtrats per eliminar continguts inadequats (Sag & Yu, 2025). Així doncs, els continguts es copien i s'incorporen a repositoris interns del desenvolupador. Tal com descriu l'EUIPO (2025), aquesta recopilació implica necessàriament la creació de còpies digitals dels materials utilitzats, fet que explica que el debat jurídic se centri inicialment en la qualificació d'aquests actes com a reproducció. La gravetat d'aquesta fase recau especialment sobre l'ús de llocs de pirateria (com LibGen o Books3) per alimentar el model (*Kadrey v. Meta* (2025)).

En aquesta fase és important el tipus de dades que s'han de recollir. Per exemple, per al processament del llenguatge natural són imprescindibles dades textuais. En aquest

sentit, la sentència de Bartz v. Anthropic (2025) evidencia com els llibres són una font fonamental per l'entrenament de la IA ja que són obres meticulosament editades, sense errors ortogràfics ni gramaticals, amb fets contrastats i ben escrits. Així mateix, en el litigi s'exposa que aquesta fase implica la còpia de milions de llibres i imatges per treballar sobre elles en un conjunt d'entrenament. En aquest mateix sentit, el tribunal del cas Kadrey v. Meta (2025) es va pronunciar sobre els llibres per ser especialment valuosos per l'entrenament; tot i ser llargs, tenen un mateix estil i coherència interna. A més, per al desenvolupament del model *Llama* descriu que van requerir l'ús de fonts diverses com Common Crawl, Wikipedia, GitHub, Project Gutenberg i bases de dades de llibres com Books3. Aquestes fonts proporcionen el volum i la diversitat necessaris perquè el model pugui generar text en múltiples estils i contextos.

2n. Pre-processament i construcció de conjunts de dades

Un cop recopilades les dades, es realitza el pre-processament. Aquesta fase consisteix a netejar, eliminar duplicats, etiquetar, filtrar i estandarditzar el material. Aquestes operacions permeten transformar un conjunt heterogeni de textos o imatges en un conjunt de dades estructurat i utilitzable per l'algoritme (Li et al., 2024). En el cas de Bartz v. Anthropic (2025), el tribunal constata que cada obra era sotmesa a un procés de depuració per eliminar elements de menor rellevància o repetitius, com ara capçaleres, peus de pàgina o numeració, obtenint així una versió netejada del contingut. Així mateix, si un mateix llibre apareixia duplicat o si, en revisar el conjunt provisional de dades, es detectava algun motiu per excloure determinades obres o categories, l'empresa disposava de la capacitat per eliminar aquestes còpies del conjunt d'entrenament.

Així mateix, l'EUIPO (2025) explica que els conjunts de dades derivats de Common Crawl o d'altres fonts no s'utilitzen en brut, sinó que acostumen a passar per processos de curació i filtratge per eliminar soroll, continguts inadequats, duplicats o materials de baixa qualitat. Aquesta etapa és decisiva perquè condiciona quins patrons aprendrà el model i redueix alguns riscos tècnics i jurídics associats a l'entrenament. Les activitats que es realitzen són la neteja, l'estandardització, l'anotació i l'extracció de característiques de les dades recopilades (Li et al., 2024). En el cas alemany *Kneschke vs. LAION* (2024), el *Landgericht* d'Hamburg descriu com l'organització va extreure URLs d'imatges del conjunt Common Crawl, va descarregar els arxius, va verificar les descripcions textuais i va crear un conjunt de dades optimitzat per entrenar sistemes d'IA d'imatge.

Des del punt de vista tècnic, aquesta etapa és imprescindible perquè el model pugui identificar patrons de manera eficient. Des del punt de vista jurídic, és rellevant perquè implica noves transformacions de còpies existents. A més, no es limiten a generar còpies merament transitòries, sinó que sovint necessiten conservar les dades durant períodes prolongats que poden comportar una reproducció de caràcter permanent de les obres utilitzades (Li et al., 2024).

3r. Tokenització i representació numèrica

La fase següent consisteix a convertir el llenguatge natural en unitats que es puguin processar pel sistema. Tal com explica el tribunal en *Kadrey v. Meta* (2025), els models de llenguatge treballen amb “*tokens*”, és a dir, unitats que poden correspondre a paraules, fragments de paraules o signes de puntuació. Cada *token* es representa mitjançant un valor numèric. En aquest sentit, el tribunal de *Bartz v. Anthropic* (2025) afirma que les còpies *tokenitzades* resultants es van utilitzar de manera reiterada durant l’entrenament. Segons algunes explicacions, aquest procés implica una dinàmica d’assaig i error mitjançant la qual el model identifica relacions estadístiques entre fragments de paraules, tant dins d’una mateixa obra com entre un volum massiu de fragments procedents d’altres llibres, pàgines web i fonts similars.

Aquesta transformació implica que el model no treballa amb textos en sentit literal, sinó amb representacions matemàtiques. No obstant, Sag i Yu (2025) remarquen que la conversió del text en *tokens* no elimina la rellevància jurídica de les còpies prèvies, atès que aquestes representacions són potencialment reversibles.

4t. Entrenament del model

Un cop les dades d’entrenament han estat degudament depurades i transformades en unitats que es poden processar, ja poden utilitzar-se per entrenar el model d’IA. Durant l’entrenament, el model analitza seqüències de paraules (*tokens*) i aprèn a predir quina és la següent més probable. Segons el cas *Bartz v Anthropic* (2025), a partir de cada intent, en base a un procés d’assaig i error, el sistema compara els resultats i ajusta progressivament els seus paràmetres, de manera que, després de moltes iteracions, millora les seves prediccions fins a obtenir un model entrenat. El model identifica relacions estadístiques entre fragments de paraules que poden ser tant dins d’una mateixa obra, com a través de grans volums de textos (EUIPO, 2025; Sag & Yu, 2025).

5è. Post-entrenament i mitigació de memorització

Un cop finalitzat l’entrenament, els models d’IAG són sotmesos a processos d’ajust addicional orientats a millorar el seu rendiment i reduir riscos associats a la reproducció de continguts protegits. En aquest sentit, el denominat *fine-tuning* permet especialitzar un model generalista mitjançant conjunts de dades més reduïts i de major qualitat, adaptant-lo a tasques concretes. Paral·lelament, tècniques com el *reinforcement learning* (que sovint implica intervenció humana) ajusten les respostes del sistema a partir de mecanismes de recompensa, alineant-les amb determinats criteris o preferències (EUIPO, 2025; Sag & Yu, 2025).

Aquests models no estan dissenyats per memoritzar literalment les dades d’entrenament, sinó per aprendre patrons i relacions estadístiques de manera abstracta. Atès que la seva capacitat és inferior al volum de dades utilitzades, el sistema opera mitjançant generalitzacions, la qual cosa redueix, en principi, el risc de reproducció exacta d’obres concretes. Ara bé, el risc de memorització no desapareix completament, especialment en supòsits en què determinats continguts són duplicats nombroses vegades o especialment distintius. Per aquest motiu, en la fase de post-entrenament s’incorporen diverses tècniques de mitigació. Entre aquestes destaquen els sistemes de

fingerprinting, que permeten detectar i evitar la generació de continguts excessivament similars a obres existents (Li et al., 2024; Sag & Yu, 2025). En la sentència de *Kadrey v. Meta* (2024), el tribunal descriu que Meta va aplicar mesures per evitar que el model memoritzés fragments textuais i els reproduís literalment. Aquestes mesures inclouen proves amb *adversarial prompting* per detectar possibles reproduccions substancials.

Amb el desenvolupament de tècniques com l'*adversarial learning* i *self-supervised learning*, la IAG ja no es limita a dades d'entrenament estàtiques, sinó que pot evolucionar mitjançant processos d'auto-iteració i optimització contínua. Això permet ampliar els límits del coneixement del model i potenciar-ne la capacitat innovadora (Li et al., 2024).

3.3. OUTPUTS GENERATS, TIPOLOGIA I RELLEVÀNCIA PER AL RISC D'INFRACCIÓ

Els *outputs* generats pels sistemes d'IA generativa suposen la manifestació externa del resultat del procés d'entrenament. Des d'un punt de vista tècnic, quan un usuari introdueix una instrucció (*prompt*), el model no accedeix a un document concret prèviament emmagatzemat ni recupera una obra identificable, sinó que genera la resposta mitjançant un càlcul de probabilitats basat en els patrons estadístics apresos durant l'entrenament (EUIPO, 2025; Fernandes, 2024). Això implica que cada *output* és, en principi, el resultat d'una combinació nova produïda en temps real a partir dels paràmetres interns del model (Fernandes, 2024; Li et al., 2024). En aquest sentit, el terme *output* fa referència al resultat generat per la IA a partir del processament d'*inputs*, el qual pot adoptar formes diverses, incloent-hi textos, imatges o música (Cernadas Lázare et al., 2024).

Tanmateix, el fet que el funcionament del model sigui probabilístic no exclou la possibilitat que el contingut generat presenti coincidències amb obres preexistents, especialment quan aquestes han format part de manera significativa del conjunt de dades d'entrenament. La probabilitat de coincidència pot augmentar en relació amb continguts repetits, àmpliament difosos o amb estructures lingüístiques molt específiques (EUIPO, 2025; Li et al., 2024). El risc d'infracció no és uniforme, sinó que és possible distingir diversos escenaris que s'estudien a l'apartat 5.

4. L'ENTRENAMENT AMB OBRES PROTEGIDES PER DRET D'AUTOR EN EL DUE

4.1. L'ENTRENAMENT DE MODELS COM A POSSIBLE ACTE DE REPRODUCCIÓ

4.1.1. Reproducció

Un dels principals interrogants jurídics que planteja l'entrenament dels sistemes d'IA és si l'ús d'obres protegides durant aquest procés pot constituir un acte de reproducció en el sentit de l'art. 2 de la Directiva InfoSoc. En aquest context, cal determinar si les operacions tècniques pròpies de l'entrenament —com la recopilació, digitalització, emmagatzematge i processament de dades— poden qualificar-se com a reproduccions jurídicament rellevants d'acord amb la interpretació desenvolupada pel TJUE.

El dret de reproducció es configura, d'acord amb l'art. 2 de la Directiva InfoSoc, com el dret exclusiu dels autors a autoritzar o prohibir “la reproducció directa o indirecta, provisional o permanent, per qualsevol mitjà i en qualsevol forma, de la totalitat o part de les seves obres”. Aquest concepte ha estat interpretat pel TJUE com un concepte autònom i uniforme del DUE, que no pot quedar sotmès a interpretacions divergents dels ordenaments nacionals (STJUE, 16 de juliol de 2009, Infopaq International, C-5/08). Es tracta d'un litigi entre l'empresa danesa Infopaq i l'associació d'editors de diaris (DDF), arran d'un sistema automatitzat que escanejava articles de premsa, els convertia en text i en generava extractes per elaborar resums per als clients, sense autorització dels titulars dels drets d'autor. Davant el conflicte, el tribunal suprem danès planteja una qüestió prejudicial al TJUE per determinar si aquestes operacions constitueixen actes de reproducció. El Tribunal estableix que qualsevol fixació d'una obra pot constituir reproducció si incorpora elements que reflecteixin l'originalitat de l'autor, fins i tot quan es tracta de fragments molt breus, com onze paraules, sempre que aquests expressin la creació intel·lectual pròpia de l'autor.

A partir d'aquesta interpretació àmplia, el criteri central per determinar l'existència de reproducció es desplaça cap a la noció de fixació de l'obra en un suport, amb independència de la seva forma tècnica o del seu grau d'accessibilitat. Així, la jurisprudència del TJUE ha reconegut que processos tècnics com l'escaneig i la conversió de continguts impliquen actes de reproducció. En particular, en el cas Infopaq, el Tribunal afirma que el procediment de tractament de dades comprèn dues activitats de reproducció: la creació d'arxius d'imatge a partir de l'escaneig d'articles de diaris impresos i la seva conversió en arxius de text.

En l'àmbit específic de la IA, tot i l'escassetat de pronunciaments del TJUE, alguns tribunals europeus han començat a abordar aquesta problemàtica. En aquest sentit, la sentència del *Landgericht* d'Hamburg de 27 de setembre de 2024 constitueix una de les primeres resolucions rellevants. El cas es fonamentava en l'activitat de LAION, una organització de recerca sense ànim de lucre que havia recopilat milers de milions d'imatges, incloent-hi obres protegides, per crear un conjunt de dades destinat a l'entrenament de models d'IA. Tot i que el tribunal no centra la seva decisió exclusivament en la qualificació jurídica d'aquestes operacions, la resolució parteix del

pressupòsit que el procés de creació del conjunt de dades implica la còpia, indexació i emmagatzematge d'obres protegides, és a dir, una forma de fixació tècnica susceptible d'encaixar en el concepte de reproducció.

En la mateixa línia, la sentència del *Landgericht* de Munic d'11 de novembre de 2025 aporta un pas més en la concreció d'aquesta qüestió. En aquest cas, la societat de gestió de drets d'autor GEMA va demandar al desenvolupador d'OpenAI arran de la generació de fragments de lletres de cançons protegides. El tribunal parteix de la premissa que aquestes obres havien estat utilitzades en l'entrenament i que, en conseqüència, havien quedat memoritzades dins del model. A partir d'aquesta constatació, afirma que aquesta memorització constitueix una forma de reproducció en el sentit del dret d'autor alemany, que transposa l'art. 2 de la Directiva InfoSoc, en la mesura que implica la fixació de l'obra dins del sistema informàtic, encara que aquesta fixació es materialitzi en forma de paràmetres o representacions internes i no en arxius identificables. En aquest sentit, el tribunal considera suficient que fragments de l'obra es trobin representats en els paràmetres del model, entenent que aquesta forma de fixació compleix els requisits per qualificar-se com a reproducció, encara que sigui tècnicament difusa.

Tot i que la jurisprudència del TJUE ha definit àmpliament el concepte de reproducció, encara existeixen pocs pronunciaments judicials que analitzin específicament l'entrenament de sistemes de IAG. Per aquest motiu, resulta especialment rellevant examinar resolucions recents d'altres tribunals internacionals que han abordat aquesta qüestió de manera directa.

En aquest context, destaca la sentència del *High Court of Justice of England and Wales* de 4 de novembre de 2025 sobre el cas *Getty Images v. Stability AI*. En aquest procediment, *Getty Images* al·legava que el model d'IA *Stable Diffusion* constituïa una còpia infractora de les imatges protegides utilitzades durant el seu entrenament, argumentant que el model es crea a partir de l'exposició a aquestes obres. No obstant, el tribunal adopta una aproximació més restrictiva, en considerar que els paràmetres del model no constitueixen una còpia de les obres utilitzades durant l'entrenament, sinó el resultat d'un procés d'aprenentatge basat en patrons i característiques abstractes. Des d'aquesta perspectiva, el model no emmagatzema les obres en si mateixes, sinó una representació estadística derivada d'aquestes, la qual cosa impediria qualificar-lo com una reproducció en sentit estricte.

En canvi, altres resolucions nord-americanes han reconegut amb més claredat l'existència d'actes de reproducció en les fases prèvies a l'entrenament. En el cas *Bartz v. Anthropic* (*U.S. District Court for the Northern District of California*, 23 de juny de 2025), el tribunal considera que la digitalització massiva de llibres (mitjançant el seu escaneig i conversió en arxius digitals) constitueix una reproducció de les obres, especialment quan aquestes còpies són emmagatzemades de manera persistent en una biblioteca digital.

De manera similar, en el cas *Kadrey v. Meta Platforms, Inc.* (*U.S. District Court for the Northern District of California*, de 25 de juny de 2025), destaca que la creació de còpies digitals completes d'obres protegides i la seva integració en conjunts de dades destinats

a l'entrenament implica la fixació de les obres en un nou suport, la qual cosa pot ser qualificada com a reproducció des de la perspectiva del dret d'autor.

Aquestes diferències jurisprudencials es reflecteixen també en l'àmbit doctrinal. Per una banda, alguns experts argumenten que no es pot considerar que hi hagi una reproducció completa o parcial de les obres originals, ja que el model no emmagatzema les dades com a còpies literals. Els *tokens* generats dins del model són una forma comprimida i abstracta de les dades d'entrenament, així doncs, consideren que per poder considerar com a reproducció segons el dret d'autor, l'obra ha de ser perceptible de manera directa o indirecta pels sentits humans. Com que a l'interior dels models d'IA no es pot accedir directament a les dades d'entrenament, sostenen que no es pot considerar que hi hagi una reproducció (Dornis, 2025; EUIPO, 2025).

D'altra banda, diversos autors sostenen que l'entrenament de sistemes de IAG implica necessàriament actes de reproducció, atès que requereix la recopilació, còpia i conservació de grans volums de dades protegides en els conjunts de dades utilitzats per entrenar els models (Li et al., 2024). En la mateixa línia, Alonso i Lucchi (2025) assenyalen que la creació de conjunts de dades mitjançant tècniques com el *web scraping* comporta la reproducció i emmagatzematge d'obres protegides, amb independència de si aquestes són posteriorment recognoscibles en els *outputs* generats.

Per la seva banda, Dornis (2025) defensa una concepció funcional del dret de reproducció, segons la qual pot existir reproducció encara que l'obra sigui transformada tècnicament, comprimida o representada de manera indirecta, sempre que es mantingui fixada en un suport des del qual pugui arribar a ser perceptible. L'autor posa com a exemple la conversió d'imatges en miniatures digitals (*thumbnails*), que la jurisprudència considera reproduccions encara que la imatge original hagi estat reduïda o comprimida. Aquest enfocament resulta especialment rellevant en el context de la IAG, on les obres són transformades en representacions estadístiques que, en determinades circumstàncies, poden permetre la reconstrucció total o parcial del contingut original mitjançant una indicació adequada en el *prompt*.

4.1.2. Aplicabilitat de l'excepció de reproduccions provisionals

El dret de reproducció no opera de manera absoluta, durant el procés d'entrenament es generen sovint còpies temporals o instrumentals, la qüestió jurídica rellevant és determinar si aquestes còpies poden quedar emparades per l'excepció prevista a l'art. 5.1 de la Directiva InfoSoc.

L'art. 5.1 estableix que els actes de reproducció temporal que siguin transitoris o accessoris, que formin part integrant i essencial d'un procés tecnològic i que tinguin com a única finalitat permetre una transmissió en una xarxa entre tercers per un intermediari o un ús lícit d'una obra, queden exempts d'autorització sempre que no tinguin una significació econòmica independent. El precepte estableix una excepció obligatòria al dret de reproducció amb l'objectiu de garantir el correcte funcionament dels processos digitals, com ara la navegació per internet o la transmissió de dades en xarxa. Per exemple, s'utilitza per la memòria RAM o la memòria cau (Iglesias Posse, 2017). Per

determinar la seva aplicabilitat al procés d'entrenament de la IAG s'han d'analitzar els requisits un per un.

Pel que fa al requisit de temporalitat, el TJUE ha establert que una reproducció només pot considerar-se transitòria quan la seva durada es limita al temps estrictament necessari per al funcionament del procés tecnològic i, especialment, quan la seva eliminació es produeix de manera automàtica i sense intervenció humana (STJUE de 16 de juliol de 2009, Infopaq International, C-5/08; STJUE de 4 d'octubre de 2011, Football Association Premier League, C-403/08 i C-429/08). No obstant, com ja s'ha comprovat en l'apartat anterior, en el cas de l'entrenament de la IAG les dades utilitzades habitualment s'emmagatzemen de manera persistent en conjunts de dades i es reutilitzen en múltiples parts del procés. Aquesta persistència és contradictòria amb el requisit de temporalitat.

En aquest sentit, el *Landgericht* d'Hamburg (2024) analitza el cas de les còpies efectuades per LAION i conclou que no s'eliminen automàticament, sinó mitjançant un procés específicament programat per a aquesta finalitat, la qual cosa implica una actuació deliberada i, per tant, no completament independent de la voluntat humana.

A més, tal com destaca la doctrina, la naturalesa transitòria de les còpies està estretament vinculada al seu caràcter volàtil, especialment en el cas de les còpies en memòria interna (RAM), que desapareixen automàticament en finalitzar el procés o interrompre's el sistema. Per contra, les còpies emmagatzemades en memòria externa, com ocorre en els conjunts de dades d'entrenament, no tenen aquest caràcter efímer, sinó que són estables i accessibles, fet que impedeix qualificar-les com a transitòries (Iglesias Posse, 2017).

Quant al caràcter accessori, la STJUE de 5 de juny de 2014, C-360/13, es pronuncia arrel del servei de seguiment de mitjans (*Meltwater*) que proporciona informes d'articles de premsa publicats a Internet, el demandant al·legava que la consulta d'un lloc web genera còpies en la pantalla de l'ordinador i en la memòria cau del disc dur. El tribunal interpreta que una reproducció és accessòria quan no té existència ni finalitat autònoma, sinó que constitueix un element purament instrumental dins d'un procés tecnològic. En aquest sentit, la doctrina ha subratllat que les reproduccions provisionals típiques són merament auxiliars, com les generades durant la transmissió de dades en xarxa, ja que no permeten l'accés ni l'explotació de l'obra i només serveixen per fer possible el funcionament tècnic del sistema. Aquesta idea es resumeix en el principi segons el qual les reproduccions instrumentals dins d'una cadena de comunicació no són rellevants en si mateixes, sinó només en relació amb el resultat final.

Tanmateix, aquesta lògica no és extrapolable a l'entrenament de sistemes d'IA. Les còpies de les obres no constitueixen un element accessori, sinó que són l'objecte mateix del procés d'anàlisi i aprenentatge (Dornis, 2025). Com ha assenyalat el *Landgericht* d'Hamburg (2024), les còpies de les imatges realitzades per la demandada es descarreguen i es copien precisament per analitzar-les, és a dir, la còpia no és un simple pas tècnic dins del procés, sinó que és l'objecte mateix de l'anàlisi.

En aquest mateix sentit s'ha pronunciat la sentència *Thomson Reuters v. Ross Intelligence* (2025). Ross va utilitzar materials derivats de opinions jurídiques protegides per entrenar el seu sistema d'IA, mitjançant la creació i ús de bases de dades intermèdies que incorporaven aquests continguts. El tribunal conclou que aquesta utilització implica una reproducció jurídicament rellevant i no merament instrumental, ja que els continguts protegits són utilitzats com a base per desenvolupar un producte competidor, i no simplement com a part transitòria d'un procés tècnic.

En tercer lloc, el requisit d'integració en un procés tecnològic exigeix que la reproducció formi part integrant i essencial d'un procés tecnològic. A diferència dels anteriors, aquest element sí que es compleix en el cas de la IAG. En efecte, l'entrenament de models d'IA implica necessàriament la ingestió, processament i anàlisi automatitzada de grans volums de dades, de manera que les reproduccions constitueixen una condició imprescindible per al funcionament del sistema.

En quart lloc es troba l'exigència de que les reproduccions tinguin com a única finalitat permetre una transmissió en xarxa entre tercers o un ús lícit de l'obra. Suposa un cas rellevant el denominat encaminament: per a que una obra sigui transmesa per la xarxa, ha de dividir-se en un grup de paquets d'informació que s'emmagatzemen de forma efímera en encaminadors i un cop arriben a l'ordinador de destí es tornen a ordenar. Com que els fragments no tenen un sentit propi, no estariem davant d'un acte de reproducció i s'aplicaria l'art. 5 de la directiva (Erdozáin López et al., 2023). Ara bé, pel que fa a l'entrenament de la IA, a diferència de l'encaminament, el procés d'entrenament implica l'emmagatzematge sistemàtic i l'anàlisi reiterada de continguts per ajustar els paràmetres del model. No es limiten a permetre la transmissió d'una obra entre tercers, sinó que constitueixen el material essencial sobre el qual es construeix el funcionament mateix del sistema, extreure informació per generar nous continguts (Contardi, 2025; EUIPO, 2025).

Per últim, l'exigència d'absència de significació econòmica independent. La doctrina ha destacat que aquest requisit es compleix en el cas de les reproduccions purament tècniques, ja que aquestes no generen valor econòmic per si mateixes ni permeten l'explotació de l'obra. Es tracta de còpies invisibles per a l'usuari i sense autonomia econòmica (Iglesias Posse, 2017). No obstant, en la IAG les còpies realitzades durant l'entrenament formen part d'un procés que té un elevat valor econòmic, ja que constitueixen la base per al desenvolupament de models comercialitzables.

Aquesta idea es veu reforçada per la sentència *Thomson Reuters v. Ross Intelligence* (2025), en què el tribunal destaca que l'ús dels continguts protegits no es limita a una funció merament tècnica o transitòria, sinó que està orientat a la creació d'un producte competidor en el mercat de serveis jurídics. Encara que les còpies no siguin explotades de manera individual, la seva integració en el procés d'entrenament contribueix directament al valor econòmic global del sistema.

En conseqüència, partint de la interpretació restrictiva que el TJUE ha fet de l'art. 5.1, resulta difícil considerar que les reproduccions realitzades durant l'entrenament de sistemes d'IA puguin quedar generalment cobertes per aquesta excepció.

4.2. L'EXCEPCIÓ DE MINERIA DE TEXTOS I DADES

La DDAMUD (art. 2.2) defineix el TDM com “qualsevol tècnica analítica automatitzada destinada a analitzar textos i dades en format digital per generar informació que inclogui, sense caràcter exhaustiu, patrons, tendències o correlacions”. En aquest sentit, el TDM constitueix una etapa central en el desenvolupament de sistemes d'IA, ja que és el mecanisme mitjançant el qual els models analitzen grans quantitats de continguts per ajustar els seus paràmetres i aprendre patrons estadístics (Contardi, 2025; EUIPO, 2025). Partint d'aquesta definició, cal examinar si les reproduccions efectuades durant les diferents fases d'entrenament de sistemes d'IA poden considerar-se emparades per les excepcions de TDM previstes a la DDAMUD.

4.2.1. L'excepció de TDM per a recerca científica (article 3 DDAMUD)

L'art. 3 DDAMUD estableix una excepció específica que permet determinats actes de reproducció i extracció d'obres quan aquests són realitzats per organismes de recerca o institucions de patrimoni cultural sense ànim de lucre amb la finalitat de dur a terme activitats de TDM amb objectius de recerca científica, sempre que aquestes entitats disposin d'accés lícit als materials utilitzats. Aquesta disposició pretén facilitar l'anàlisi automatitzada de grans volums d'informació en l'àmbit científic, per facilitar la investigació acadèmica i científica en entorns digitals, eliminant obstacles jurídics que podrien dificultar aquestes pràctiques (Contardi, 2025; García Vidal, 2024; Izquierdo, 2021).

Aquesta exigència implica que el TDM no pot legitimar l'ús de còpies obtingudes de manera il·lícita. Aquesta idea de licitud també s'ha posat de manifest en la jurisprudència internacional: en el cas *Bartz v. Anthropic* (2025), el Tribunal va analitzar la pràctica de l'empresa *Anthropic* de descarregar milions de llibres piratejats per crear una biblioteca destinada a l'entrenament de models d'IA. El tribunal va considerar que la descàrrega d'aquestes còpies piratejades constituïa una infracció per si mateixa, independentment de l'ús posterior que se'n fes. En aquest sentit, el jutge va afirmar que no és possible justificar la còpia il·lícita d'obres simplement defensant una finalitat de recerca o desenvolupament tecnològic, ja que permetre-ho podria perjudicar greument el mercat editorial acadèmic. Així doncs, d'acord amb aquesta interpretació, qualsevol ús d'obres obtingudes de manera il·lícita quedaria fora de les excepcions relatives al TDM i constituïria una vulneració del dret de reproducció.

Un dels elements centrals de l'art. 3 és la limitació subjectiva dels beneficiaris de l'excepció. Pel que fa als organismes de recerca, la Directiva inclou una àmplia varietat d'entitats que tenen com a objectiu principal la investigació científica o combinar aquesta activitat amb serveis educatius. Dins d'aquesta categoria s'hi troben, entre altres, les universitats, els instituts de recerca o altres entitats que desenvolupin activitats científiques, com ara hospitals o centres especialitzats (Izquierdo, 2021).

Tanmateix, la Directiva estableix determinades condicions perquè una entitat pugui considerar-se organisme de recerca. En particular, aquestes organitzacions han d'operar sense ànim de lucre o reinvertir els seus beneficis en activitats de recerca científica, o bé actuar en el marc d'una missió d'interès públic reconeguda per un Estat membre. Ara bé, s'exigeix que cap empresa que exerceixi una influència significativa sobre l'organització pugui gaudir d'un accés privilegiat als resultats de la investigació (Contardi, 2025). El segon grup de beneficiaris el constitueixen les institucions de patrimoni cultural, definides per la Directiva com qualsevol biblioteca o museu accessible al públic, arxiu o institució dedicada a la conservació del patrimoni cinematogràfic o sonor (Contardi, 2025; Izquierdo, 2021).

Pel que fa a les condicions d'ús, el requisit d'accés lícit té un paper central. Aquest pot derivar tant de continguts disponibles sota polítiques d'accés obert com de continguts als quals s'accedeix mitjançant acords contractuals amb els titulars de drets, com ara subscripcions a bases de dades o plataformes digitals (DDAMUD, considerant 14) (Contardi, 2025). En el primer cas, l'accés obert es refereix a informació científica disponible gratuïtament per als usuaris i susceptible de reutilització. En el segon cas, l'accés lícit inclou continguts disponibles en línia o mitjançant subscripció institucional. En aquest sentit, quan una institució de recerca o patrimoni cultural disposa d'una subscripció, també es considera que tenen accés lícit les persones vinculades a aquesta institució i que operen sota aquesta cobertura (Izquierdo, 2021).

A més, la norma també preveu que les còpies generades durant aquestes activitats puguin emmagatzemar-se sota condicions adequades de seguretat i conservar-se durant el temps necessari per a la investigació científica, especialment amb l'objectiu de verificar o reproduir els resultats obtinguts. Al mateix temps, es reconeix als titulars de drets la possibilitat d'adoptar mesures destinades a garantir la seguretat i integritat de les xarxes i bases de dades on es troben allotjades les obres o prestacions. No obstant, aquestes mesures no poden excedir el que sigui estrictament necessari per assolir aquesta finalitat, de manera que no es converteixin en un obstacle desproporcionat per a l'excepció (considerant 16, DDAMUD) (García Vidal, 2024).

Malgrat aquest intent d'equilibri entre els interessos dels titulars de drets i les necessitats de la recerca científica, part de la doctrina ha advertit que les limitacions associades al requisit d'accés lícit i a les mesures tecnològiques de protecció poden restringir l'abast pràctic de l'excepció i situar la UE en una posició menys competitiva ja que pot suposar que l'entrenament es faci amb continguts més bàsics i no actualitzats (Contardi, 2025).

En últim lloc, una de les característiques essencials de l'excepció és el seu caràcter obligatori i no renunciable, que implica la prohibició de limitar o anular l'excepció mitjançant acords contractuals o del mecanisme d'*opt-out*, recollit a l'art. 7 DDAMUD. Aquesta característica respon a l'objectiu del legislador europeu de garantir la seguretat jurídica i facilitar el desenvolupament de la recerca científica basada en l'anàlisi massiva de dades (Alonso & Lucchi, 2025; Serranía, 2020).

Un cop analitzats els diferents elements que configuren l'art. 3, és convenient valorar l'encaix de l'entrenament de la IAG. La jurisprudència europea recent ha començat a abordar aquesta qüestió, en concret la sentència del *Landgericht* d'Hamburg (2024). El tribunal alemany va considerar aplicable l'excepció de TDM per a recerca científica a les activitats de l'organització LAION, ja que es tracta d'una entitat sense ànim de lucre dedicada a la creació de conjunts de dades per a la investigació en IA. El tribunal va entendre que LAION podia qualificar-se com a organisme de recerca donat que la seva activitat consistia en desenvolupar i posar a disposició de la comunitat científica conjunts de dades i coneixement sobre l'entrenament d'algoritmes d'IA. En particular, el tribunal va considerar que la creació d'un conjunt de dades destinat a l'entrenament de models d'IA podia constituir una activitat de TDM amb finalitats científiques, ja que implicava l'anàlisi automatitzada de continguts per extreure informació sobre la relació entre imatges i descripcions textuais. En concret, el tribunal considera que la descàrrega de la imatge objecte del litigi per comparar-la amb una descripció d'aquesta mateixa imatge és una tècnica analítica automatitzada destinada a generar una correlació que encaixa perfectament amb la definició de TDM.

El tribunal utilitza la regla dels tres passos, recollida a l'art. 5.4 de la Directiva Infosoc i aplicable a la TDM per disposició expressa de l'art. 7.2 DDAMUD, i considera que l'ús de la fotografia, limitat a la seva anàlisi automatitzada per establir correspondències amb descripcions textuais, no substitueix ni interfereix amb el mercat de l'obra original, ja que no afecta la seva explotació econòmica. Tot i reconèixer que els conjunts de dades resultants poden ser utilitzats per entrenar sistemes d'IA capaços de generar continguts potencialment competitiu amb les obres humanes, el tribunal entén que aquest risc és hipotètic i no constitueix, per si sol, un perjudici actual i injustificat. En concret, assenyalava que acceptar aquest argument implicaria prohibir qualsevol activitat de TDM, la qual cosa seria contrària a la voluntat expressa del legislador europeu.

Ara bé, la mateixa sentència introdueix una distinció rellevant entre tres fases diferents del desenvolupament de sistemes d'IA: la creació del conjunt de dades d'entrenament, el posterior entrenament del model neuronal, i la utilització del sistema entrenat per generar nous continguts. Segons el tribunal, l'activitat analitzada en el cas era la creació del conjunt de dades que sí podia quedar coberta per l'excepció de TDM. Encara que no es va pronunciar de manera definitiva sobre si l'entrenament dels models d'IA en si mateix queda també emparat per aquesta excepció.

Des d'una perspectiva doctrinal, divergeixen les opinions sobre l'aplicabilitat del TDM. Una part rellevant de la doctrina sosté que l'excepció de TDM ha de ser aplicable a l'entrenament de sistemes d'IAG. Segons aquesta posició, igual que en les tècniques de TDM, l'entrenament de models generatius es limitaria a l'anàlisi d'informació semàntica, de manera que els elements expressius protegits pel dret d'autor no serien objecte d'explotació directa. Així, defensen que el procés d'entrenament no implica una apropiació de l'expressió, sinó una extracció de patrons i relacions subjacents a les dades.

Per una altra part de la doctrina, cal distingir entre el TDM i el funcionament dels models generatius. Mentre que el TDM té com a objectiu principal analitzar grans volums de

dades per identificar patrons, relacions i informació rellevant, els sistemes d'IAG utilitzen aquestes dades com a base del seu procés d'aprenentatge. Així, el model no només capta regularitats estadístiques, sinó també les formes d'expressió contingudes en els exemples d'entrenament, que li permeten generar nous continguts. En conseqüència, a diferència del TDM, la finalitat d'aquests sistemes no és l'extracció d'informació, sinó la producció de resultats nous a partir dels patrons apresos (Ordelin Font, 2025; Rosati, 2025).

En aquesta mateixa postura, s'argumenta que l'aplicació de l'excepció de TDM a l'entrenament d'IAG entra en conflicte amb els principis bàsics del dret d'autor europeu. A diferència del TDM, l'entrenament dels models generatius no es limita a analitzar informació abstracta, sinó que implica la incorporació integral dels continguts utilitzats, incloent-hi els elements expressius protegits de les obres. En conseqüència, aquesta posició sosté que l'excepció de TDM no resulta aplicable, ni tampoc cap altra excepció prevista en el dret d'autor (Dornis, 2025).

Per tant, de l'anàlisi conjunta de la doctrina i de la jurisprudència es pot concloure que l'excepció de l'art. 3 pot resultar aplicable a determinades fases prèvies al desenvolupament de sistemes d'IA, com ara la recopilació i anàlisi de dades per crear conjunts de dades amb finalitats de recerca científica. Tanmateix, encara no existeix un consens clar sobre si l'entrenament dels models generatius pròpiament dit pot considerar-se també una activitat de TDM.

4.2.2. L'excepció general de TDM (article 4 DDAMUD)

L'art. 4 DDAMUD introdueix una excepció general de TDM que permet la realització de reproduccions d'obres protegides, sempre que aquestes siguin lícitament accessibles i que la finalitat sigui l'anàlisi automatitzada d'informació. A diferència de l'art. 3, aquesta excepció no es limita a determinats subjectes ni a finalitats de recerca científica, sinó que pot ser utilitzada per qualsevol persona, ja sigui física o jurídica, pública o privada, per activitats desenvolupades amb finalitats comercials (EUIPO, 2025; Navas Navarro, 2025; Vela, 2024). Aquesta amplitud respon a la voluntat del legislador europeu de fomentar la innovació, especialment en sectors privats (Ordelin Font, 2025). Alhora, contribueix a garantir unes condicions de competència equitatives en el mercat europeu, exigint que tots els proveïdors (independentment del lloc on s'hagi realitzat l'entrenament) respectin els estàndards del dret d'autor de la UE (Li et al., 2024). En aquest sentit, Navas Navarro (2025) assenyala que l'art. 4 DDAMUD és l'excepció a la qual s'han d'emparar les empreses que desenvolupen models d'IAG amb finalitats comercials.

No obstant, l'art. 4 presenta la possibilitat que els titulars de drets reservin expressament l'ús de les seves obres per tal de limitar aquest TDM de propòsit general. Aquest mecanisme, conegut com a *opt-out*, permet als titulars oposar-se a l'ús de les seves obres per a TDM mitjançant mitjans adequats. Això implica que l'excepció només opera en absència d'aquesta reserva (EUIPO, 2025; Li et al., 2024; Navas Navarro, 2025; Serranía, 2020).

Pel que fa als requisits materials, tant l'art. 3 com l'art. 4 exigeixen que l'accés a les obres sigui lícit, de manera que queden exclosos els continguts obtinguts de fonts il·legals (EUIPO, 2025; Serranía, 2020). Així mateix, l'abast de l'excepció es limita al dret de reproducció i bases de dades, sense estendre's a altres drets d'explotació com la transformació o la comunicació pública (Navas Navarro, 2025). En aquest sentit, l'excepció cobreix únicament les operacions necessàries per a la realització de TDM, però no els usos posteriors que puguin derivar-se del tractament de les dades (Contardi, 2025; Vela, 2024). Una altra diferència rellevant respecte de l'art. 3 es troba en la durada de la conservació de les còpies, l'art. 4 estableix que aquestes només poden mantenir-se durant el temps estrictament necessari per a la realització de TDM. Això reforça el caràcter instrumental i limitat d'aquesta excepció (García Vidal, 2024; Serranía, 2020).

Per tal de comprovar l'aplicabilitat en l'entrenament d'IAG, hem de recórrer de nou a la sentència del *Landgericht* d'Hamburg (2024). El tribunal clarifica que la descàrrega d'imatges per a la creació d'un conjunt de dades d'entrenament d'IA es pot subsumir en el concepte de TDM previst a l'art. 4 DDAMUD (transposat a l'art. 44b UrhG de la legislació alemanya). Ara bé, la resolució acaba desestimant l'aplicació del precepte a causa de l'existència d'una reserva de drets vàlida per part del titular.

En el cas *Authors Guild v. Google* (2015), els tribunals nord-americans van considerar lícita la digitalització massiva d'obres protegides amb la finalitat de crear un sistema de cerca, malgrat que implicava la còpia íntegra dels llibres. El tribunal va entendre que les obres eren utilitzades per permetre l'anàlisi i indexació del seu contingut. Aquesta lògica és propera al TDM, en la mesura que l'objectiu és extreure informació i detectar patrons, i no explotar l'obra com a tal. A més, el cas encaixaria dins l'excepció de l'art. 4 atès que Google és una corporació comercial amb ànim de lucre (Izquierdo, 2021; Vercelli, 2022).

Segons Stober i Dornis (2026), el TDM té com a objectiu extreure informació i descobrir patrons a partir de grans volums de dades, mentre que els models generatius són entrenats per reproduir la distribució probabilística de les dades d'entrenament amb la finalitat de generar nous continguts similars als originals. L'excepció de TDM pot resultar aplicable a determinades fases preparatòries del desenvolupament de sistemes d'IA, especialment aquelles vinculades a la recopilació, selecció, neteja, estructuració i *tokenització* de dades destinades a la creació de conjunts de dades d'entrenament. No obstant, existeixen dubtes jurídics sòlids sobre la seva extensió a les fases d'entrenament, post-entrenament i generació de resultats, atès que aquestes activitats ja no persegueixen una finalitat estrictament analítica, sinó generativa (Alonso & Lucchi, 2025; Contardi, 2025).

4.2.3. Reserva de drets: mecanisme d'opt-out, requisits i efectivitat pràctica

Un dels elements centrals de l'excepció de TDM prevista a l'art. 4 DDAMUD és el mecanisme de reserva de drets (*opt-out*), que determina directament l'àmbit d'aplicació (Ordelin Font, 2025). Encara que l'accés lícit a les obres constitueix un requisit previ indispensable per a l'aplicació de l'excepció, els titulars de drets tenen la facultat

d'excloure les seves obres del TDM mitjançant una reserva expressa, d'acord amb l'art. 4.3 de la Directiva (Navas Navarro, 2025).

Aquesta reserva ha de realitzar-se de manera adequada: mitjançant mitjans de lectura mecànica en el cas de continguts disponibles en línia, com ara metadades o les condicions d'ús d'un lloc web (art. 4.3 DDAMUD; considerant 18). Així mateix, la reserva també pot articular-se per altres vies, com acords contractuals o declaracions unilaterals dels titulars (Izquierdo, 2021). En la pràctica, això es tradueix en mecanismes com protocols d'exclusió de robots o sistemes destinats a impedir el *web scraping*, que permeten als titulars controlar l'accés automatitzat als seus continguts (EUIPO, 2025; Navas Navarro, 2025). En conseqüència, quan existeix una reserva vàlida, l'ús de les obres per a finalitats de TDM comercial requereix l'autorització prèvia dels titulars.

Des d'una perspectiva sistemàtica, la validesa d'un *opt-out* exigeix tres requisits: que la reserva sigui expressa, que sigui efectuada pel titular dels drets i que es realitzi mitjançant mitjans adequats. Pel que fa al primer, existeixen dues posicions: una d'estricta, que exigeix una referència clara a la TDM, i una altra de més àmplia, que admet reserves generals com prohibicions de *web scraping*. Igualment, els titulars poden reservar els seus drets a través de les condicions generals del lloc web, sempre que aquestes siguin degudament acceptades pels usuaris (Navas Navarro, 2025). Quant al requisit que la reserva sigui feta pel titular dels drets, poden sorgir dificultats en contextos amb múltiples titulars o intermediaris, tot i que en alguns casos aquesta facultat pot ser exercida per llicenciataris o entitats de gestió col·lectiva. Finalment, respecte als mitjans adequats, es considera que en entorns digitals la reserva ha de ser comprensible per sistemes automatitzats, encara que la pràctica i la jurisprudència apunten a una interpretació flexible que també pot incloure condicions d'ús en llenguatge natural (EUIPO, 2025). Així doncs, la interpretació concreta d'aquests requisits presenta encara incerteses i podria evolucionar a través de la jurisprudència.

La sentència del *Landgericht* d'Hamburg (2024) aporta un primer pronunciament rellevant sobre la interpretació pràctica del mecanisme d'*opt-out*. El tribunal assenyalava que la reserva de drets pot considerar-se vàlida encara que no adopti una forma tècnica especialment sofisticada, sempre que resulti clarament expressada i accessible. En concret, el tribunal suggereix que les condicions generals d'un lloc web poden complir el requisit de reserva adequada, fins i tot si estan redactades en llenguatge natural, la qual cosa apunta cap a una interpretació flexible del criteri de mitjans de lectura mecànica.

A més, la decisió posa de manifest que el mecanisme d'*opt-out* no exigeix necessàriament una referència explícita de TDM, sinó que podria ser suficient una prohibició general d'extracció automatitzada o *scraping*. Això té una conseqüència important: amplia potencialment l'àmbit de les reserves de drets i reforça la posició dels titulars, ja que facilita que puguin excloure les seves obres del TDM sense necessitat d'utilitzar fórmules jurídiques específiques. Aplicat al cas concret, el tribunal considera que els indicis de reserva existents podrien complir els requisits de l'art. 4.3, tot i que no fixa un criteri definitiu, deixant oberta la qüestió a futures interpretacions jurisprudencials.

Adicionalment, en aquest context, el Reglament d'IA reforça la importància pràctica del mecanisme de reserva de drets, tot i no regular directament el TDM, en exigir als proveïdors de models d'IA que respectin les reserves expressades pels titulars. Això converteix l'*opt-out* en un instrument clau per delimitar la licitud de l'entrenament de sistemes d'IAG en el marc europeu (Ordelin Font, 2025).

Tanmateix, la doctrina ha assenyalat també les limitacions d'aquest model. D'una banda, el sistema de reserva de drets s'aproxima funcionalment a un model de consentiment previ, doncs, un cop exercida la reserva els desenvolupadors han d'obtenir autorització per utilitzar les obres (Ordelin Font, 2025). D'altra banda, aquest esquema pot generar costos de transacció elevats (tant d'identificació dels titulars com de negociació de llicències) i dificultar l'accés a grans volums de dades, imprescindibles per a l'entrenament (EUIPO, 2025). A més, pot donar lloc a biaixos en els models, en limitar la diversitat de les dades d'entrenament, amb possibles implicacions en termes d'eficàcia dels sistemes (Contardi, 2025; Li et al., 2024).

4.3. TRANSPARÈNCIA I COMPLIMENT NORMATIU EN RELACIÓ AMB CONTINGUTS D'ENTRENAMENT

La irrupció dels sistemes d'IAG ha posat de manifest la dificultat dels titulars de drets per conèixer si les seves obres han estat utilitzades en els processos d'entrenament dels models. Molts casos es topen amb una barrera probatòria pel que fa a l'acreditació de l'ús de material protegit per drets d'autor en l'entrenament d'IA, atès que els *outputs*, tot i estar influenciats pels conjunts de dades emprats, no solen reproduir íntegrament les obres, la qual cosa obliga els titulars a fonamentar les seves pretensions en similituds detectades en els resultats com a prova indirecta (Ayo Ferrándiz et al., 2025; Contardi, 2025). Aquesta manca de transparència no només genera incertesa, sinó que incideix directament en la possibilitat d'exercir els drets reconeguts per l'ordenament jurídic, especialment en relació a l'aplicabilitat de les excepcions. En absència de coneixement per part dels titulars, la mera facultat d'exercir la reserva de l'art. 4 esdevé insuficient com a mecanisme per garantir l'efectiva protecció dels drets d'autor (Alonso & Lucchi, 2025; Contardi, 2025; Li et al., 2024; Stober & Dornis, 2026).

Davant d'aquest escenari, el Reglament d'IA incorpora diverses obligacions específiques en matèria de transparència adreçades als proveïdors de models d'IA de propòsit general, en particular, l'art. 53 estableix dues obligacions rellevants en matèria de transparència i compliment normatiu. D'una banda, l'art. 53.1.c disposa que els proveïdors han de disposar d'una política interna destinada a garantir el respecte dels drets d'autor, incloent-hi mecanismes adequats per identificar i complir les reserves d'exclusió formulades pels titulars dels drets d'acord amb l'art. 4.3 DDAMUD. D'altra banda, l'art. 53.1.d imposa l'obligació d'elaborar i posar a disposició del públic un resum suficientment detallat del contingut utilitzat per a l'entrenament dels models, d'acord amb la plantilla que proporciona l'Oficina d'IA, sense comprometre secrets empresarials ni informació comercial confidencial (Contardi, 2025; EUIPO, 2025; Fernandes, 2024; Li et al., 2024; Sag & Yu, 2025).

Ara bé, aquest marc normatiu presenta limitacions importants, tant des d'un punt de vista conceptual com pràctic. En primer lloc, la desconexió temporal i conceptual entre la DDAMUD i el Reglament d'IA: mentre l'excepció de TDM opera de forma *ex ante*, l'obligació de transparència opera *ex post*, referint-se a dades que ja han estat utilitzades pel model. Aquesta divergència temporal fa que la transparència actuï principalment com a mecanisme de verificació posterior i no com una eina preventiva suficient (Ayo Ferrándiz et al., 2025; Contardi, 2025; Li et al., 2024).

En segon lloc, la implementació pràctica de les obligacions de transparència es veu dificultada per la naturalesa massiva i heterogènia de les dades d'entrenament (Fernandes, 2024; Li et al., 2024). La inexistència d'un registre obligatori d'obres, la fragmentació de titularitats i la manca d'estandardització de les metadades fan extremadament complex identificar totes les obres incorporades als conjunts de dades (Contardi, 2025; Fernandes, 2024). En aquest sentit, el considerant 107 estableix que el resum de les dades d'entrenament ha de ser suficientment ampli per permetre l'exercici dels drets, però no ha de ser tècnicament exhaustiu ni identificar cada obra individualment (Fernandes, 2024; Li et al., 2024). Així mateix, diversos autors consideren que una transparència exhaustiva seria, en molts casos, materialment inviable o excessivament costosa (Contardi, 2025; Ordellin Font, 2025). A més, el Reglament d'IA contempla l'aplicació del principi de proporcionalitat, especialment per a pimes i *start-ups*, tot i que la seva aplicació pràctica continua sent incerta (Fernandes, 2024; EUIPO, 2025).

Pel que fa al contingut del resum, la normativa apunta com a exemple la identificació de les principals col·leccions o conjunts de dades utilitzats per a l'entrenament, juntament amb una explicació narrativa d'altres fonts emprades. A més, Cotino Hueso (2023) assenyalava que ha d'incloure informació sobre l'origen de les dades, les categories utilitzades, els criteris de selecció i altres aspectes rellevants per comprendre el funcionament del sistema (Li et al., 2024). Aquesta informació és essencial no només per a la protecció dels drets d'autor, sinó també per a la detecció de biaixos i la garantia de la qualitat dels sistemes. Així, la transparència té una funció dual: protegir els titulars i reforçar la confiança i la responsabilitat en l'ús de la IA (Ayo Ferrándiz et al., 2025; Cotino Hueso, 2023). Tanmateix, el Reglament d'IA també limita l'abast de la transparència per protegir secrets comercials i informació confidencial, cosa que pot reduir la utilitat dels resums per als titulars de drets (EUIPO, 2025; Fernandes, 2024).

En tercer lloc, les obligacions de transparència poden esdevenir merament formals si no s'acompanyen de mecanismes efectius de control i execució. Segons aquesta visió, la publicació de polítiques de compliment o de resums sobre dades d'entrenament podria convertir-se en un requisit burocràtic sense impacte real en la protecció dels titulars (Alonso & Lucchi, 2025).

Malgrat aquestes limitacions, el Reglament d'IA té un paper rellevant en la configuració dels deures de diligència dels operadors d'IA. En establir obligacions de documentació, actualització i divulgació d'informació sobre les dades utilitzades, el Reglament contribueix a definir estàndards de comportament que poden ser utilitzats per valorar el

compliment normatiu dels proveïdors. D'aquesta manera, la transparència no només protegeix els titulars de drets, sinó que també s'integra en un marc més ampli de compliment normatiu i gestió de riscos dels sistemes d'IA (Ayo Ferrándiz et al., 2025; Cotino Hueso, 2023; Li et al., 2024). A més, des d'un punt de vista tècnic i ètic, promoure la documentació estructurada fomenta la responsabilitat i redueix els riscos de memorització no desitjada de dades d'entrenament (Stober & Dornis, 2026).

Aquest enfocament es veu reforçat pel caràcter extraterritorial del Reglament d'IA, que obliga a complir les seves disposicions a tots els proveïdors que introdueixin models al mercat de la Unió, independentment del lloc on s'hagin desenvolupat o entrenat (Fernandes, 2024; Li et al., 2024). Això pretén evitar avantatges competitius derivats de l'ús de jurisdiccions amb estàndards de protecció més baixos i garantir un nivell homogeni de protecció dins del mercat europeu (EUIPO, 2025; Fernandes, 2024). No obstant, l'efectivitat real d'aquesta extraterritorialitat dependrà de la capacitat d'execució i dels possibles conflictes de dret internacional privat (Sag & Yu, 2025).

En definitiva, la transparència és necessària, però no suficient. El Reglament d'IA representa un pas important perquè introdueix obligacions específiques de publicitat i compliment que poden millorar la posició jurídica dels titulars i reduir, en part, l'opacitat del procés d'entrenament (EUIPO, 2025; Fernandes, 2024). No obstant, els seus límits estructurals, probatoris i pràctics impedeixen afirmar que, per si sola, garanteixi una protecció efectiva dels drets d'autor. Més aviat, la transparència apareix com una condició necessària per a l'exercici dels drets, però insuficient si no va acompanyada d'una normativa més precisa que permeti transformar aquesta informació en una tutela realment efectiva (Alonso & Lucchi, 2025; Ordelin Font, 2025).

5. LA DIMENSIÓ PROBATÒRIA DELS *OUTPUTS*: LA GENERACIÓ DE CONTINGUTS COM A INDICI DE REPRODUCCIÓ EN LA FASE D'ENTRENAMENT

Els *outputs* generats pels sistemes d'IA no només plantegen un possible risc autònom d'infracció, sinó que també poden operar com a indici probatori del que ha succeït prèviament en la fase d'entrenament. En efecte, si el resultat generat reproduceix totalment o parcialment una obra protegida, aquest *output* pot revelar que l'obra no només ha estat utilitzada com a dada d'entrada, sinó també assimilada pel model en una forma suficientment estable per ser exterioritzada posteriorment. Així, la generació de continguts pot funcionar com una externalització dels *inputs* (Dornis, 2025; EUIPO, 2025).

Des d'una perspectiva jurídica, la qüestió fonamental és determinar si aquests *outputs* constitueixen una reproducció il·lícita, segons l'art. 2 Directiva Infosoc i la jurisprudència del TJUE, la qual cosa exigeix analitzar si el resultat incorpora elements originals recognoscibles de l'obra preexistent i si aquesta similitud es produeix en el pla de l'expressió protegida o únicament en el de les idees. Atenent al grau de similitud substancial entre l'obra generada i l'obra preexistent, podem categoritzar els *outputs* en tres tipologies amb diferents implicacions probatòries sobre la vulneració del dret de reproducció (Fernandes, 2024; Rosati, 2025).

5.1. LA REGURGITACIÓ LITERAL O MEMORITZACIÓ: L'*OUTPUT* COM A PROVA IRREFUTABLE DE LA FIXACIÓ DE L'OBRA AL MODEL

El supòsit de risc més intens es produeix quan el sistema genera un *output* idèntic o substancialment idèntic a una obra protegida emprada en l'entrenament. Aquesta situació, coneguda com a memorització o regurgitació, no és simplement una coincidència funcional del sistema, sinó un indici especialment fort que el model ha conservat informació expressiva de l'obra en els seus paràmetres. Si un model és capaç de generar un fragment extens d'un llibre, una cançó o una imatge protegida de manera pràcticament literal, resulta difícil sostenir que l'obra només ha estat analitzada de manera abstracta. Com assenyala Dornis (2025), la reaparició total o parcial d'una obra protegida en l'*output* implica que aquesta obra es troba representada dins del model d'alguna manera jurídicament rellevant. Aquesta tesi és coherent amb els estudis tècnics recents que indiquen que la probabilitat de regurgitació augmenta amb la mida del model, la freqüència del contingut en el conjunt de dades i la longitud o especificitat del *prompt* (EUIPO, 2025; Liu et al., 2026). Per tant, la reproducció no pot considerar-se accidental quan es tracta de fragments extensos o complexos, ja que la probabilitat d'una coincidència espontània és extremadament reduïda.

La regurgitació literal encaixa amb el concepte de reproducció de l'art. 2 Directiva InfoSoc, ja que suposa la fixació i posterior exteriorització d'una obra o d'una part substancial d'aquesta. Des del punt de vista jurisprudencial, la sentència del *Landgericht* de Munic de 11 de novembre de 2025 estableix que la memorització de lletres de cançons en models de llenguatge constitueix una vulneració del dret de reproducció. La

resolució assevera que les obres es troben fixades físicament al model i, consegüentment, el fet que la IA les faci perceptibles a l'usuari a través dels seus resultats infringeix el dret de reproducció. A més, encara que l'empresa presenti la regurgitació com un error, una anomalia o un ús abusiu per part de l'usuari, l'existència d'un *output* literal revela que el sistema contenia una reproducció derivada del procés d'entrenament (Dornis, 2025; EUIPO, 2025; Liu et al., 2026).

Pel que fa a les excepcions de TDM dels arts. 3 i 4 DDAMUD, determinats actes d'extracció i reproducció en la fase d'anàlisi o entrenament poden quedar coberts, però no autoritzen la generació posterior d'un *output* que reproduïx una obra protegida atès que aquesta no és la seva finalitat (Navas Navarro, 2025; Rosati, 2025).

5.2. LA INCORPORACIÓ D'ELEMENTS CREATIUS RECONOSCIBLES COM A PROVA DE LA REPRODUCCIÓ PARCIAL

El segon supòsit és més complex: l'*output* no reproduïx íntegrament l'obra, però incorpora elements creatius recognoscibles d'aquesta. L'anàlisi no ha de preguntar-se si l'*output* és semblant en termes generals, sinó si ha incorporat una part protegida de l'obra preexistent. En aquest sentit, cal recordar que l'art. 2 Directiva InfoSoc prohibeix expressament la reproducció tant "total com parcial". D'acord amb la STJUE *Infopaq* (2009), la reproducció parcial és jurídicament rellevant quan el fragment reproduït expressa l'originalitat de l'autor, encara que sigui quantitativament reduït. La STJUE de 4 de desembre de 2025 (afers C-580/23 i C-795/23) afegeix que per declarar una infracció del dret de reproducció, no s'ha d'atendre a la impressió visual general, sinó determinar si s'han incorporat de forma recognoscible objectivament els elements creatius que reflecteixen les decisions lliures i creatives de l'autor.

Aquest criteri és especialment important perquè evita dues conclusions extremes. D'una banda, impedeix afirmar que tot *output* derivat d'un model entrenat amb obres protegides és necessàriament infractor, sinó que és necessari un anàlisi cas per cas. En *Kadrey v. Meta* (2025), el tribunal va rebutjar la pretensió massa abstracta que tots els *outputs* de *Llama* serien obres derivades, però va precisar que els demandants haurien d'al·legar i provar que els *outputs* incorporaven alguna porció dels llibres protegits i que existia similitud substancial. D'altra banda, també impedeix reduir la infracció als casos de còpia literal. Samuelson (2023) adverteix que no n'hi ha prou amb demostrar que el resultat està basat en obres d'entrenament, sinó que cal acreditar l'apropiació d'una quantitat substancial d'expressió original; tanmateix, també reconeix que els *outputs* poden infringir quan reproduïxen elements protegits, especialment en supòsits de duplicació de dades, manca de filtres o *prompts* dirigits a obtenir contingut específic.

La STS 1755/2023, cas Kukuxumusu, també resulta útil per delimitar aquesta frontera. El Tribunal Suprem admet que pot haver-hi reproducció total o parcial quan un nou dibuix no sigui realment diferent i pugui qualificar-se de plagi, però rebutja una prohibició general de redibuixar personatges en escenes diferents si el resultat és efectivament distint. Aplicat als *outputs* d'IA, això implica que no tota aparició d'un personatge, temàtica o estructura similar constitueix reproducció; només ho serà quan l'*output*

incorpori la forma expressiva concreta i recognoscible de l'obra anterior (Samuelson, 2023).

A més, la STJUE de 14 d'abril de 2026 (cas *pastiche*) en matèria de *sampling* (la reutilització de fragments d'obres preexistents) ha establert que fins i tot la incorporació de fragments breus d'una obra a una altra pot constituir infracció quan aquests són recognoscibles per l'oient o l'espectador.

Des del punt de vista del TDM, podria plantejar-se si aquests *outputs* són una simple conseqüència indirecta del procés analític i, per tant, haurien de beneficiar-se de l'excepció. Tanmateix, aquesta interpretació resulta difícil de sostenir: la DDAMUD defineix la TDM com una tècnica destinada a analitzar informació per extreure'n coneixement, no a reproduir o recrear continguts protegits. Per això, la doctrina insisteix que l'entrenament de models generatius va més enllà del TDM, ja que no només analitza, sinó que internalitza i pot regurgitar contingut en forma d'*outputs* (Contardi, 2025; García Vidal, 2024). Això implica que, encara que la incorporació parcial d'elements pugui tenir origen en el procés d'entrenament, l'acte jurídicament rellevant és el resultat generat, no l'anàlisi prèvia. I aquest resultat, si incorpora expressió protegida, no pot considerar-se un simple ús analític (Rosati, 2025).

5.3. ANÀLISI DE LA REPLICACIÓ ESTILÍSTICA COM A FRONTERA DE LA FACULTAT DE REPRODUCCIÓ

Finalment, una tercera categoria està formada pels *outputs* que no reproduïen cap element concret d'una obra protegida, sinó que es limiten a evocar l'estil, el to, les idees, el gènere, la tècnica o determinades característiques estructurals associades a un autor o corrent creatiu, sense incorporar fragments identificables d'una obra concreta (EUIPO, 2025; Sag & Yu, 2025).

Si un model d'IA és capaç de generar una obra nova que reproduïx únicament l'estil d'un autor, això *per se* no serveix com a prova concloent que s'hagi infringit el dret de reproducció durant l'entrenament, ja que el model hauria extret patrons estadístics que no es protegeixen (Samuelson, 2023). La major part de la doctrina considera que no existeix infracció del dret d'autor, atès que l'estil no és un element protegit per la PI. Aquests resultats solen quedar fora de l'àmbit de protecció del copyright perquè l'estil, així com les idees o els conceptes generals, no són objecte d'apropiació exclusiva (Navas Navarro, 2025; Sag & Yu, 2025). Així, encara que un *output* pugui recordar l'obra d'un autor determinat, això no és suficient per apreciar una vulneració si no es produeix una apropiació de la seva expressió concreta (Samuelson, 2023). En la jurisprudència europea, la recent STJUE de 4 de desembre de 2025, assumptes acumulats C-580/23 i C-795/23, recorda “el principi fonamental dels drets d'autor segons el qual no es protegeixen les idees, sinó únicament les seves expressions”. En aquest sentit, en el cas *Kadrey v. Meta* (2025), el tribunal conclou que “*Meta volia que el model Llama fos capaç de generar text en determinats estils. Però l'estil no és objecte de protecció per drets d'autor: només ho és l'expressió*”. Així mateix, el tribunal del cas *Bartz v. Anthropic* (2025)

recorda que el *copyright* no protegeixen els mètodes de funcionament, ni les idees o principis que una obra pugui mostrar o en què es basi.

Tanmateix, cal advertir, com demostren Liu et al. (2026), que l'emulació d'estil pot ser el detonant que activi accidentalment la recuperació de fragments memoritzats textualment, convertint el que semblava una imitació d'estil lícita en una reproducció parcial il·lícita. Sota l'aparença d'estar imitant un estil, el model reproduceix accidentalment seqüències protegides de les obres originals, un fenomen que la recerca ha demostrat que ocorre amb freqüència quan es realitza *fine-tuning* sobre un autor específic.

6. CONCLUSIONS

I. L'objectiu principal d'aquest treball era determinar si l'entrenament de sistemes d'IA pot constituir una infracció de la facultat de reproducció en el dret d'autor de la UE. A partir de l'anàlisi normativa, jurisprudencial i doctrinal realitzada, concloc que, amb caràcter general, sí que es produeixen actes de reproducció jurídicament rellevants durant l'entrenament dels models. L'afirmació es fonamenta en el fet que el procés d'entrenament exigeix necessàriament la còpia, fixació i processament de continguts protegits perquè el model pugui aprendre patrons estadístics i estructures expressives. Per tant, el funcionament dels sistemes generatius no pot desvincular-se completament dels continguts originals utilitzats durant l'entrenament.

II. Tot i això, considero que no totes les fases tècniques del procés presenten el mateix nivell de rellevància jurídica. La principal controvèrsia es troba en la *tokenització*. La fragmentació de les obres en unitats mínimes pot dificultar la identificació d'una expressió protegida, especialment quan els fragments tractats de manera aïllada deixen de reflectir elements originals de l'obra. Tanmateix, diversos autors han provat que és possible recuperar les reproduccions originals. Per aquest motiu la *tokenització* constitueix el límit més discutible de la reproducció dins l'entrenament dels models d'IA.

III. Pel que fa a la hipòtesi relativa a si les reproduccions generades durant l'entrenament podien quedar emparades per l'excepció de reproduccions provisionals de l'art. 5.1 de la Directiva Infosoc, concloc que, amb caràcter general, aquesta excepció no resulta aplicable a l'entrenament de sistemes generatius. La meua conclusió deriva de l'anàlisi acumulativa dels requisits exigits pel TJUE. Les còpies produïdes durant l'entrenament no són merament transitòries ni accessòries, sinó estructurals i indispensables per al funcionament mateix del sistema. A més, tampoc considero que manquin de valor econòmic independent, ja que el valor comercial dels models depèn precisament de les dades utilitzades durant l'entrenament. Per aquest motiu, s'entén que intentar encaixar l'entrenament dins l'art. 5.1 implica estendre excessivament una excepció concebuda originalment per a processos tècnics molt diferents.

IV. Respecte del TDM, l'art. 4 DDAMUD constitueix la principal cobertura jurídica de què disposen actualment els desenvolupadors de sistemes d'IA dins de la UE, atesa la seva finalitat comercial. Tanmateix, aquesta excepció exigeix com a requisit essencial que les dades utilitzades hagin estat obtingudes de manera lícita i el respecte a les reserves realitzades pels titulars de drets. A més, la recerca m'ha portat a concloure que el TDM i l'entrenament de models generatius no són activitats plenament equivalents. La principal raó és que el TDM està concebut per permetre l'extracció automatitzada d'informació, patrons o correlacions, mentre que els models generatius no es limiten a analitzar dades, sinó que poden reproduir estructures expressives i generar continguts similars a les obres entrenades. Per aquest motiu considero que l'excepció de TDM pot justificar bona part de les reproduccions necessàries durant la fase prèvia d'entrenament, però no resol completament els problemes jurídics associats als *outputs* generatius ni cobreix íntegrament totes les fases tècniques del desenvolupament dels models.

V. Quant al mecanisme d'*opt-out* de l'art. 4.3 DDAMUD, aquest actua com una eina de control que permet als titulars excloure les seves obres de l'entrenament de la IA. Actualment sembla suficient una reserva general de drets expressada mitjançant mecanismes amplis sense necessitat d'una formulació tècnica específica, sempre que resulti clarament expressada i accessible. Tot i això, la seva eficàcia pràctica es veu limitada per la manca d'estàndards universals i els elevats costos de negociació que poden limitar la qualitat de les dades emprades en l'entrenament.

VI. En relació amb la transparència i les obligacions introduïdes pel Reglament d'IA, la transparència constitueix un requisit imprescindible per garantir una protecció efectiva dels titulars de drets. Sense informació mínima sobre les dades utilitzades durant l'entrenament, els mecanismes d'*opt-out* i les possibilitats de control resulten pràcticament ineficaços. Tot i això, considero que la regulació actual continua sent insuficient, tant per la manca de mecanismes realment eficaços de control i verificació, com per les dificultats inherents a la naturalesa massiva, heterogènia i dinàmica de les dades utilitzades. A més, una transparència excessivament detallada podria afectar secrets comercials, incrementar de manera significativa els costos tècnics del desenvolupament i fins i tot incentivar l'ús de dades més limitades, obsoletes o de menor qualitat. Per això considero que encara no s'ha assolit una solució que permeti un equilibri entre la protecció dels titulars de drets i el foment de la innovació.

VII. Pel que fa als *outputs* generats pels sistemes d'IA, és imprescindible distingir entre diferents categories de resultats. Quan el sistema genera una reproducció literal o gairebé idèntica d'una obra entrenada, considero que existeix una infracció clara del dret de reproducció, ja que l'*output* exterioritza directament contingut original i l'expressió creativa de l'autor. En canvi, quan la reproducció és parcial, s'ha d'atendre a una valoració cas per cas de la importància qualitativa dels elements reproduïts i de si aquests reflecteixen decisions creatives recognoscibles de l'autor. Així, aquests *outputs* infractors no poden quedar emparats per l'excepció de TDM, ja que aquesta només cobreix l'extracció i anàlisi de dades durant l'entrenament, i no la posterior generació de continguts. Finalment, la mera imitació d'un estil no constitueix, en general, una infracció del dret de reproducció, pel fet que el dret d'autor només protegeix expressions concretes i no idees, tècniques o estètiques generals.

VIII. En resum, la irrupció de la IAG ha suposat un repte important per al dret d'autor, atès que obliga a replantejar l'equilibri entre la protecció dels drets de propietat intel·lectual i la promoció de la innovació tecnològica. Moltes de les qüestions analitzades, especialment les relatives a l'abast del dret de reproducció, a l'aplicabilitat de les excepcions de TDM, a l'efectivitat del mecanisme d'*opt-out* i les obligacions de transparència, encara no han estat resoltes definitivament pels tribunals europeus. Aquesta manca de jurisprudència i interpretació consolidada, juntament amb la ràpida evolució tecnològica de la matèria, evidencia la necessitat de continuar desenvolupant futures investigacions i d'analitzar la progressiva concreció normativa i jurisprudencial d'aquest àmbit.

7. LLISTAT DE REFERÈNCIES

7.1. LLISTAT DE REFERÈNCIES

Alonso, E., & Lucchi, N. (2025). *AI And Copyright* (SSRN Scholarly Paper Núm. 5401059). Social Science Research Network. <https://papers.ssrn.com/abstract=5401059>

Ayo Ferrándiz, C., Seijo Bar, Á., Garre Anguera de Sojo, I., & González Guillén, P. (2025). *RESPONSABILIDAD CIVIL E INTELIGENCIA ARTIFICIAL*.

Bragado Herrero de Egaña, C. L. (2023). *Los retos de la propiedad intelectual en el entorno digital*. Dykinson. <https://elibro.net/es/lc/uab/titulos/251906>

Candelario Macias, I., Rojas Romero, E., Medina González, M. ángel, Dopazo Fraguío, P., Blas A. González, Juan Garbayo Blanch, Ruiz Muñoz, M., Francesco Mattina, Isabel Ramos Herranz, Isabel Blanco Esguevillas, Mónica Lastiri Santiago, Juan José Caselles Fornés, Sara Martín Salamanca, Raquel Sampedro Calle, Vega Justribó, B. de la, & Teixeira Rodríguez, M. (2017). *Derecho de la Propiedad Intelectual*. <https://biblioteca-tirant-com.eu1.proxy.openathens.net/cloudLibrary/ebook/info/9788491693567>

Cernadas Lázare, M., Torres Pérez, F. J., Fernández Carballo Calero, P. I., Tato Plaza, A., Louredo Casado, S., & Costas Comesaña, J. (2024). *Nuevas tendencias en el derecho de la competencia y de la propiedad industrial e intelectual IV*. 723-747.

Contardi, M. (2025). El rol de la excepción de minerías de datos en la inteligencia artificial (IA) generativa – perspectiva europea. *Revista e-Mercatoria*, 24(1). <https://doi.org/10.18601/16923960.v24n1.02>

Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida. *Revista Española de la Transparencia*, (16), 17-63. <https://doi.org/10.51915/ret.272>

Díaz Noci, J. (2023). Inteligencia artificial, noticias y medios de comunicación: Una aproximación jurídica desde la perspectiva de la propiedad intelectual al concepto y atribución de autoría. *Textual & Visual Media: revista de la Sociedad Española de Periodística*. 2023;17(1):7-21. <https://doi.org/10.56418/txt.17.1.2023.1>

Dornis, T. W. (2025). Generative AI, Reproductions Inside the Model, and the Making Available to the Public. *IIC - International Review of Intellectual Property and Competition Law*, 56(5), 909-938. <https://doi.org/10.1007/s40319-025-01582-9>

Erdozain López, J. C., Minero Alejandro, G., Bercovitz Álvarez, G., Juan José Marin López, Bercovitz Rodríguez-Cano, R., Garrote Fernández Díez, I., Cámara Águila, M. del P., González Gozalo, A., Moralejo Imbernón, N., Sánchez Aristi, R., López Maza, S., Morillo González, F., & Martínez Espin, P. (2023). *Manual de Propiedad Intelectual 10ª Edición*. <https://biblioteca-tirant-com.eu1.proxy.openathens.net/cloudLibrary/ebook/info/9788411694469>

EUIPO. (2025). *The development of generative artificial intelligence from a copyright perspective*. Publications Office. <https://doi.org/10.2814/3893780>

Fernandes, P. M. (2024). AI Training and Copyright: Should Intellectual Property Law Allow Machines to Learn? *Bioethica*, 10(2), 8-21. <https://doi.org/10.12681/bioeth.39041>

García Vidal, Á. (2024). *Propiedad intelectual, minería de textos y datos y entrenamiento de la inteligencia artificial*.

Iglesias Posse, R. (2017). Las reproducciones provisionales de obras protegidas en el ámbito digital. Especial referencia a las generadas durante el intercambio de archivos através de redes P2P. Segunda parte. *Dereito: revista xurídica da Universidade de Santiago de Compostela*, 26(1). <https://doi.org/10.15304/dereito.26.1.3672>

Izquierdo, H. A. (2021). Minería de textos y datos e Inteligencia Artificial: Nuevas excepciones al derecho de autor. *THEMIS Revista de Derecho*, (79), 323-343. <https://doi.org/10.18800/themis.202101.018>

Li, K., Wu, H., & Dong, Y. (2024). Copyright protection during the training stage of generative AI: Industry-oriented U.S. law, rights-oriented EU law, and fair remuneration rights for generative AI training under the UN's international governance regime for AI. *Computer Law & Security Review*, 55, 106056. <https://doi.org/10.1016/j.clsr.2024.106056>

Liu, X., Miresghallah, N., Ginsburg, J. C., & Chakrabarty, T. (2026). *Alignment Whack-a-Mole: Finetuning Activates Verbatim Recall of Copyrighted Books in Large Language Models* (arXiv:2603.20957). arXiv. <https://doi.org/10.48550/arXiv.2603.20957>

Montaño, B. B. (2022). La inspiración en la inteligencia artificial y el caso de los datos cuyo contenido son obras protegidas por el derecho de autor. *Revista de Derecho*, 21(41), 33-52. <https://doi.org/10.47274/DERUM/41.3>

Navas Navarro, S. (2025). The training of AI models in the context of the EU Copyright law and the AI act. *Open Journal of Social Science*, 13, 263-277. <https://doi.org/10.4236/jss.2025.138016>

Ordelin Font, J. L. (2025). Derechos de autor y entrenamiento de sistemas de IA generativos: Las obligaciones de transparencia y la minería de textos y datos en la normativa europea. *IDP. Revista de Internet, Derecho y Política*, (42), 1-13. <https://doi.org/10.7238/idp.v0i42.431327>

Organización Mundial de la Propiedad Intelectual (Ed.). (2020). *¿Qué es la propiedad intelectual?* World Intellectual Property Organization (WIPO). <https://doi.org/10.34667/tind.44180>

Rogel Vide, C., & Serrano Gómez, E. (2021). *Manual de Derecho de autor*. <https://elibro.net/es/ereader/uab/269319?prev=as>

Rosati, E. (2024). Is text and data mining synonymous with AI training? *Journal of Intellectual Property Law and Practice*, 19(12), 851-852. <https://doi.org/10.1093/jiplp/jpae092>

Rosati, E. (2025). Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law. *European Journal of Risk Regulation*, 16(2), 603-627. <https://doi.org/10.1017/err.2024.72>

Sag, M., & Yu, P. K. (2025). The Globalization of Copyright Exceptions for Ai Training. *Emory Law Journal*, 74(5), 1163-1227.

Samuelson, P. (2023). Generative AI meets copyright. *Science*, 381(6654), 158-161. <https://doi.org/10.1126/science.adi0656>

Serranía, V. J. (2020). Datos, minería e innovación: Qvo vadis, Europa? Análisis sobre las nuevas excepciones para la minería de textos y datos. *CUADERNOS DE DERECHO TRANSNACIONAL*, 12(1), 247-258. <https://doi.org/10.20318/cdt.2020.5188>

Stober, S., & Dornis, T. W. (2026). *Generative AI Training and Copyright Law* (arXiv:2502.15858). arXiv. <https://doi.org/10.48550/arXiv.2502.15858>

Vela, J. M. M. (2024). Inteligencia artificial generativa. Desafíos para la propiedad intelectual. *Revista de Derecho de la UNED (RDUNED)*, (33), 17-75. <https://doi.org/10.5944/rduned.33.2024.41924>

Vercelli, A. (2022). Copyright, derecho de autor e inteligencias artificiales: Análisis de sus procesos de adecuación a los cambios tecnológicos. *JAIIO, Jornadas Argentinas de Informática*, 8(11), 35-46.

7.2. JURISPRUDÈNCIA

High Court of Justice (Chancery Division). Sentència de 4 de novembre de 2025, *Getty Images (US) Inc. v. Stability AI Ltd*

Landgericht Hamburg. Sentència de 27 de setembre de 2024, assumpte 310 O 227/23, *Kneschke / LAION e.V.*

Landgericht München I. Sentència de 11 de novembre de 2025, assumpte 42 O 14139/24, *GEMA / Suno AI GmbH*

Superior Court of California, County of San Francisco, *Reddit, Inc. v. Anthropic, PBC*, Demanda presentada el 4 de juny de 2025.

Tribunal de Justícia de la Unió Europea (Gran Sala) Sentència de 14 d'abril de 2026, assumpte C-590/23, *CG i YN / Pelham GmbH i altres*

Tribunal de Justícia de la Unió Europea (Gran Sala). Sentència de 4 d'octubre de 2011, assumptes acumulats C-403/08 i C-429/08, *Football Association Premier League Ltd i altres / QC Leisure i altres*, ECLI:EU:C:2011:631

Tribunal de Justícia de la Unió Europea (Sala Primera). Sentència de 4 de desembre de 2025, assumptes acumulats C-580/23 i C-795/23,

Mio AB i altres / Galleri Mikael & Thomas Asplund Aktiebolag i USM U. Schärer Söhne AG / konektra GmbH, ECLI:EU:C:2025:941

Tribunal de Justícia de la Unió Europea (Sala Quarta). Sentència de 5 de juny de 2014, assumpte C-360/13, *Public Relations Consultants Association Ltd / Newspaper Licensing Agency Ltd i altres*, ECLI:EU:C:2014:1195

Tribunal de Justícia de la Unió Europea (Sala Quarta). Sentència de 16 de juliol de 2009, assumpte C-5/08, *Infopaq International A/S / Danske Dagblades Forening*, ECLI:EU:C:2009:465

Tribunal Suprem (Sala Primera, Civil). Sentència núm. 1755/2023, de 19 de desembre de 2023, rec. 905/2020, *Kukuxumusu Ideas S.L. / Íñigo i altres*, ECLI:ES:TS:2023:5648

United States Copyright Office. *Zarya of the Dawn (Registration #VAu001480196)*. Letter decision, 21 de febrer de 2023.

United States Court of Appeals (Second Circuit). *Authors Guild v. Google Inc.* (Google Books), 16 d'octubre de 2015.

United States District Court for the Northern District of California, *Concord Music Group, Inc. et al. v. Anthropic PBC*.

United States District Court for the Southern District of New York, Ordre de 4 d'abril de 2025, *The New York Times Company et al. v. OpenAI, Inc. et al. (motions to dismiss)*.

United States District Court for the Southern District of New York, *Authors Guild et al. v. OpenAI, Inc. et al.*, Demanda presentada el 19 de setembre de 2023.

United States District Court, District of Delaware. *Thomson Reuters Enterprise Centre GmbH i West Publishing Corp. v. Ross Intelligence Inc.*, No. 1:20-cv-613-SB, Memorandum Opinion, 11 de febrer de 2025.

United States District Court, Northern District of California. *Bartz v. Anthropic PBC*, No. C 24-05417 WHA, Order on Fair Use, 23 de juny de 2025.

United States District Court, Northern District of California. *Kadrey et al. v. Meta Platforms Inc.*, Case No. 3:23-cv-03417-VC, 25 de juny de 2025