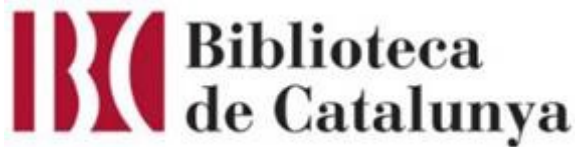


vidimus

vidimus SL
B60898137
Av. Mare de Déu de
Montserrat 207 local 3
08041 Barcelona
(+34) 609 32 66 46 | info@vidimus.net
www.vidimus.net



Informe sobre l'estat de la intel·ligència artificial (IA) a Europa en relació amb la catalogació de documents

BIBLIOTECA DE CATALUNYA

INFORME

02/12/2025

Contingut

1. Resum executiu	5
1.1. Estructura de l'informe.....	5
2. Metodologia	7
2.1. Revisió sistemàtica de fonts	7
2.1. Intel·ligència de mercat i indicadors.....	7
2.2. Estudi de camp: enquesta a institucions europees sobre l'ús de la intel·ligència artificial en projectes de catalogació de documents	7
2.2. Selecció de casos d'ús	8
3. Context. La Intel·ligència Artificial en l'ecosistema tecnològic bibliotecari	9
3.1. Percepció i extracció de dades: de l'OCR a la Visió per Computador	9
3.1.1. Aplicació pràctica i evidència empírica.....	9
3.2. Aprenentatge automàtic i classificació: de la percepció a la cognició.....	10
3.2.1. Evidència d'eficàcia i limitacions	10
3.3. Processament del Llenguatge Natural (PLN) i Models de Llenguatge (LLMs)	10
3.3.1. Aplicació en la catalogació descriptiva i enriquiment.....	10
3.3.2. La «paradoxa de l'eficiència».....	11
3.3.3. Mitigació d'al·lucinacions: arquitectura RAG	11
3.3.4. Normalització del text i morfologia	11
3.4. Reconeixement d'Entitats Nomenades (<i>Named Entity Recognition</i> - NER) i control d'autoritats	12
3.4.1. De la identificació a la reconciliació d'entitats.....	12
3.5. IA simbòlica, vectorització i grafs de coneixement	12
3.5.1. Vectorització (Embeddings) i cerca semàntica.....	12
3.5.2. Arquitectura RAG.....	13
3.5.3. Cap a les Dades Enllaçades (<i>Linked Data</i>)	13
3.6. Síntesi de tecnologies i aplicacions	13
3.7. Prospectiva tecnològica: cap a la IA neuro-simbòlica	14
4. Context institucional: adopció i marcs de referència a Europa	15
4.1. L'escenari europeu: entre l'experimentació i la producció	15
4.2. L'Observatori d'Intel·ligència Artificial de REBIUN	15

5.	Anàlisi del sector tecnològic i aplicació en processos tècnics	16
5.1.	El mercat global de tecnologia bibliotecària	16
5.1.1.	L'estratègia de la «IA Invisible»	16
5.1.2.	Consolidació i divergència tecnològica	16
5.2.	Catalogació automatitzada: de la ingesta al registre estructurat	16
5.2.1.	Ingesta de documents digitals (PDF a MARC)	16
5.2.2.	Catalogació Visual i fluxos híbrids	17
5.2.3.	Identificació de la manifestació i gestió d'adquisicions	17
5.2.4.	Limitacions empíriques i dependència normativa	17
5.3.	Implementació assistida d'RDA i model LRM	17
5.4.	Anàlisi de contingut: indexació, classificació i resum	18
5.4.1.	Indexació semàntica automatitzada: l'estàndard Annif	18
5.4.2.	Classificació automàtica: precisió i limitacions disciplinàries	18
5.4.3.	Generació de resums i enriquiment contextual	19
5.5.	Control de Qualitat i Manteniment del Catàleg (<i>Data Hygiene</i>)	19
5.5.1.	Detecció de duplicats difusos	19
5.5.2.	Normalització i neteja massiva	19
5.5.3.	Validació post-generació	20
5.6.	Interoperabilitat i Web semàntica: la transició cap a Linked Data	20
5.6.1.	El debat al voltant de les ontologies bibliotecàries: és necessària l'ontologia en l'era dels LLMs?	20
5.6.2.	Conversió intel·ligent: de cadenes de text a URI	20
6.	Desafiaments ètics i problemes tecnològics: al·lucinacions, biaixos i marc legal	21
6.1.	La problemàtica de les al·lucinacions	21
6.2.	Biaixos heretats i limitacions culturals	21
6.3.	La paradoxa de l'eficiència i el cost operatiu	22
6.4.	Incertesa jurídica i propietat intel·lectual	22
7.	Estudi de camp: enquesta a institucions europees sobre l'ús de la intel·ligència artificial en projectes de catalogació de documents	23
7.1.	Perfil dels enquestats	23
7.2.	Nivell d'adopció actual i àrees d'aplicació prioritària	25
7.2.1.	Adopció actual de la IA	26

7.2.2.	Ús i valoració de la IA: àmbits d'interès.....	27
7.2.3.	Tècniques d'IA explorades o implementades	27
7.2.4.	Tasques específiques d'implementació de la IA actualment.....	29
7.2.5.	Integració de la IA en Models de Metadades.....	30
7.3.	Enfocament específic en catalogació i la seva integració amb la IA	31
7.3.1.	Dades de catalogació anual i proporció manual.....	31
7.3.2.	Expectatives d'impacte de la IA en la catalogació	32
7.3.3.	Èxit percebut en la implementació de la IA en catalogació	33
7.4.	Factors organitzatius i desafiaments en la implementació de la Intel·ligència Artificial	34
7.4.1.	La preparació institucional: desafiaments en la planificació.....	35
7.4.2.	Perfils Clau en la Implantació: La Direcció i el Catalogador	35
8.	Resultats i consideracions finals	37
8.1.	El nou rol professional	37
8.2.	L'equilibri entre sostenibilitat i costos.....	37
8.3.	Cap a la convergència: IA neuro-simbòlica.....	37
8.4.	Recomanacions estratègiques	37
9.	Bibliografia i fonts	39

1. Resum executiu

El present informe analitza l'estat de la tecnologia bibliotecària a Europa i la implantació de la Intel·ligència Artificial (IA) en tasques vinculades amb la catalogació durant el període 2020-2025, encara que amb especial atenció en els darrers dos anys. Aquest període es caracteritza per la convergència entre els sistemes de gestió tradicionals i la irrupció operativa de la Intel·ligència Artificial (IA) Generativa. La investigació examina la transformació dels processos de catalogació i descripció documental, documentant el trànsit des de models manuals cap a ecosistemes de descripció assistits per algoritmes.

L'anàlisi del mercat revela una dicotomia estructural. Per una banda, s'observa una consolidació de l'oligopoli comercial (LSP/ILS) que integra la IA mitjançant una **estratègia «invisible»** orientada a optimitzar fluxos de treball existents. Per una altra, emergeix una avantguarda d'innovació en biblioteques nacionals i universitàries europees que, buscant la sobirania tecnològica, desenvolupen models propis i adopten arquitectures neuro-simbòliques.

Per complementar l'anàlisi teòrica i de mercat, l'informe incorpora un Estudi de camp basat en una enquesta a institucions bibliotecàries europees. Aquest estudi de camp ofereix una visió empírica sobre la realitat operativa, detallant els actuals nivells d'adopció de la IA en els fluxos de treball, les àrees d'aplicació prioritàries (com la Traducció automàtica i l'OCR) i els desafiaments organitzatius i estratègics a què s'enfronten les institucions en la implementació pràctica.

Els resultats semblen indicar que la tecnologia no està reemplaçant al perfil professional, sinó redefinint el seu rol cap a la **supervisió semàntica**. No obstant això, s'identifiquen barreres crítiques per a l'adopció massiva: l'al·lucinació inherent als models generatius, la persistència de biaixos heretats en les dades d'entrenament i la incertesa jurídica en el marc de la Unió Europea.

1.1. Estructura de l'informe

El present document s'articula en vuit capítols:

1. **Resum executiu.** El present apartat. Síntesi de les troballes clau, objectius de la investigació i full de ruta del document.
2. **Metodologia.** Descripció del disseny de la investigació, detallant la revisió sistemàtica de fonts i el disseny del treball de camp (enquesta europea).
3. **Context tecnològic.** Anàlisi de les capacitats tècniques de la IA aplicada enfront de les necessitats de la descripció bibliogràfica.
4. **Context institucional.** Examen de l'adopció i marcs estratègics a Europa contrastant l'experimentació amb la producció real.
5. **Anàlisi del sector i processos tècnics.** Avaluació del mercat de proveïdors (LSP/ILS) i desglossament de casos d'ús pràctics: catalogació original, normativa RDA, indexació automàtica i control de qualitat.
6. **Desafiaments ètics i crítics.** Identificació de barreres estructurals, incloent-hi la problemàtica de les al·lucinacions, els biaixos heretats i la incertesa legal.

7. **Estudi de camp.** Presentació i anàlisi dels resultats obtinguts en l'enquesta realitzada a institucions bibliotecàries europees sobre l'ús de la IA.
8. **Resultats i consideracions finals.** Consideracions sobre el nou rol professional, la sostenibilitat econòmica i l'orientació dels sistemes d'Intel·ligència Artificial cap a un model d'IA Neuro-simbòlica.
9. **Bibliografia i fonts.** Amb una selecció de la bibliografia i fonts utilitzades en aquest treball.

2. Metodologia

L'elaboració d'aquest estudi s'ha regit per un disseny metodològic mixt que combina l'anàlisi de la bibliografia actual al voltant del tema en qüestió, la situació del mercat tecnològic de les biblioteques, el disseny, elaboració, enviament i anàlisi d'un qüestionari per recopilar dades a nivell europeu i la revisió de dades secundàries procedents d'enquestes sectorials. Aquest enfocament permet contrastar la teoria al voltant de la IA amb l'oferta tecnològica teòrica i la realitat de les institucions.

2.1. Revisió sistemàtica de fonts

S'ha analitzat un corpus documental produït principalment entre 2020 i 2025, classificat en tres tipologies:

- **Informes tècnics.** Documentació d'arquitectura de sistemes (Annif, FOLIO) i *white papers* de proveïdors tecnològics (Ex Libris, OCLC, etc.).
- **Literatura científica.** Articles revisats per parells sobre l'aplicació de sistemes d'IA en catalogació i en processos afins, prioritzant estudis experimentals d'institucions europees.
- **Literatura Gris.** Actes de conferències especialitzades, grups d'interès especial, tesis doctorals, etc.

2.1. Intel·ligència de mercat i indicadors

L'anàlisi del sector privat es fonamenta en l'exegesi de la sèrie *Library Systems Report* (Breeding, 2023-2025) i l'*International Survey of Library Automation*. Aquests documents constitueixen la font primària per avaluar la penetració real de funcionalitats d'IA. Per al diagnòstic d'adopció institucional, s'han sintetitzat les dades de quatre instruments: l'enquesta de la Conferència de Biblioteques Nacionals Europees (CENL, 2024), l'informe de l'Observatori d'IA de REBIUN (2025), el *report global Pulse of the Library* (Clarivate, 2025) i el treball de camp que s'ha elaborat per a aquest treball.

2.2. Estudi de camp: enquesta a institucions europees sobre l'ús de la intel·ligència artificial en projectes de catalogació de documents

Per complementar l'anàlisi teòrica, es va dissenyar i desplegar un instrument d'investigació dirigit a una selecció estratègica d'institucions bibliotecàries europees. L'objectiu d'aquest estudi de camp ha estat el de validar el grau de penetració real de les tecnologies d'IA en els departaments de processos tècnics.

Per a això s'ha construït un qüestionari estructurat en set blocs temàtics, dissenyat per discriminar entre l'adopció general de la IA i la seva aplicació específica en catalogació:

1. Identificació i tipologia. Segmentació per país, tipus de biblioteca (pública, universitària, nacional) i volum de col·lecció.
2. Nivell d'adopció general. Filtre inicial per distingir entre ús actiu, exploració teòrica o nul·la implementació.
3. Tecnologies implementades. Identificació d'arquitectures específiques (PLN, Visió per Computador, *Machine Learning* i Robòtica).

4. Ecosistema d'eines. Mapeig de programari, diferenciant entre solucions comercials (ChatGPT, Alma, Copilot...) i desenvolupaments de codi obert o propis.
5. Aplicació en catalogació. Bloc central que avalua l'ús d'IA en tasques crítiques (catalogació original, assignació de matèries, control d'autoritats) i estàndards de metadades (MARC21, BIBFRAME, RDA).
6. Governança i implementació. Anàlisi dels perfils professionals involucrats i estudis previs de viabilitat.
7. Avaluació d'impacte. Valoració qualitativa sobre l'eficiència, qualitat de les dades i satisfacció del personal mitjançant escales.

Mostra i procediment. El qüestionari va ser distribuït a un directori d'institucions bibliotecàries seleccionades per la seva rellevància en l'espai europeu. Les dades obtingudes, corresponents a 87 respostes vàlides, van ser sotmeses a una anàlisi estadística descriptiva per identificar tendències, barreres comunes i correlacions entre la grandària de la institució i la seva maduresa tecnològica.

2.2. Selecció de casos d'ús

Finalment, s'ha mapejat un directori d'institucions de referència que lideren la innovació aplicada.

3. Context. La Intel·ligència Artificial en l'ecosistema tecnològic bibliotecari

La integració de la Intel·ligència Artificial (IA) a les biblioteques no constitueix un fenomen monolític, al contrari, correspon més amb una agregació de tecnologies diferents i en molts casos aïllades que impacten en diferents capes de la infraestructura bibliotecària. Per comprendre el seu abast, s'ha plantejat en aquest informe abandonar les habituals definicions generals i procedir a analitzar les capacitats tècniques específiques que aquestes eines habiliten en el context de la descripció bibliogràfica i el processament tècnic.

3.1. Percepció i extracció de dades: de l'OCR a la Visió per Computador

L'automatització de la catalogació s'inicia amb la capacitat del sistema per processar el document, especialment quan aquest és físic i no té una representació digital. En aquest domini, la tecnologia ha evolucionat des del Reconeixement Òptic de Caràcters (OCR) tradicional cap a la Visió per Computador semàntica (*Computer Vision*). Els sistemes actuals quan processen una imatge (per exemple, la imatge resultat de la digitalització d'una font per a la descripció d'un document, com la portada) no es limiten a la transcripció d'informació visual (píxels) a text (ASCII), sinó que empen xarxes neuronals convolucionals (CNN) per realitzar una «Anàlisi de Disseny de Documents» (*Document Layout Analysis - DLA*).

Aquesta capacitat tècnica permet al sistema segmentar una imatge (per exemple, la imatge digitalitzada de la portada) i distingir semànticament els elements basant-se en la topologia visual i la jerarquia tipogràfica i no exclusivament en el contingut lèxic. D'aquesta manera, l'algoritme pot inferir que un text ubicat a la part superior amb un cos de font major correspon al camp Títol, mentre que un text a la part inferior amb menor grandària correspon a l'Editorial. A més, s'obre la possibilitat que sobre la base de catalogacions prèvies i la interacció humana el sistema pugui evolucionar i «aprendre» els detalls del procés.

3.1.1. Aplicació pràctica i evidència empírica

La literatura recent ofereix casos d'ús que validen l'eficiència operativa d'aquestes tecnologies.

Un paradigma d'aplicació és el flux de treball desenvolupat per la Biblioteca Reial de Bèlgica (KBR). Lowagie (2023) detalla la implementació d'un sistema automatitzat mitjançant *Microsoft Power Automate i AI Builder*. Aquest flux ingereix imatges escanejades de portades i colofons procedents del dipòsit legal, utilitzant models de visió entrenats per detectar entitats bibliogràfiques i estructurar les dades en camps MARC preliminars. Els resultats reportats indiquen una reducció del temps de processament d'un registre bàsic a menys de 2,5 minuts, demostrant la viabilitat de la visió per computador per a la gestió massiva.

Per altra banda, a partir de les tècniques de Reconeixement de Text Manuscrit (HTR), s'han fet aplicacions pràctiques per a la gestió de fons històrics, l'HTR ha superat les barreres de l'OCR tradicional. Projectes com *NorHand*, de la Biblioteca Nacional de Noruega, o PARES (Portal de Archivos Estatales del Ministerio de Cultura) utilitzen models basats en tecnologia HTR per transcriure manuscrits amb variacions

cal·ligràfiques complexes, fent que el contingut dels arxius sigui indexable i descobrible a text complet sense necessitat d'una transcripció humana.

3.2. Aprenentatge automàtic i classificació: de la percepció a la cognició

Mentre que la Visió per Computador resol el desafiament de la percepció (llegir el document), l'Aprenentatge Automàtic (*Machine Learning* - ML) aborda la cognició classificatòria, una competència tradicionalment exclusiva de l'intel·lecte bibliotecari.

Aquesta tecnologia s'aplica fonamentalment mitjançant tècniques **d'Aprenentatge Supervisat**. En el context de la catalogació, aquest procés implica l'entrenament d'algoritmes utilitzant grans volums de registres MARC preexistents (denominats tècnicament *Ground Truth*). L'objectiu és que el sistema aprengui a associar patrons lingüístics i metadades (com títols o índexs) amb codis de classificació estandarditzats (CDU, Dewey, Llistes d'Encapçalaments de Matèria, etc.).

3.2.1. Evidència d'eficàcia i limitacions

La literatura científica confirma la viabilitat d'aquests models per a tasques de classificació massiva, encara que amb matisos importants respecte a la seva autonomia. Així, Chadwick (2022), en la seva investigació aplicada a l'*Schools Catalogue Information Service* (SCIS), va demostrar que classificadors estadístics poden alinear recursos bibliogràfics amb àrees curriculars específiques assolint una precisió superior al 90% en categories amb àmplia representació de dades, de la qual cosa deriva que es pot aconseguir una bona precisió en entorns controlats.

És crucial destacar que, a diferència de l'assignació determinista humana, la IA opera com un sistema probabilístic. L'algoritme no assigna una signatura topogràfica absoluta, sinó que suggereix una classificació acompanyada d'un nivell de confiança o aproximació probabilística (per exemple, un 85% de probabilitat). Això redefineix el flux de treball, situant l'expert humà en un rol de validació final per mitigar biaixos en categories marginals o menys representades en les dades d'entrenament.

3.3. Processament del Llenguatge Natural (PLN) i Models de Llenguatge (LLMs)

El camp del Processament del Llenguatge Natural (PLN) ha experimentat una profunda evolució en el darrer lustre. La disciplina ha transitat des d'enfocaments basats en regles rígides (com l'eliminació de *stopwords* o l'*stemming* bàsic) cap a l'adopció generalitzada de Grans Models de Llenguatge (*Large Language Models* - LLMs).

Aquesta evolució, que marca el pas de l'estadística de freqüència a la comprensió semàntica profunda (*Deep Learning*), permet als sistemes actuals analitzar no només la presència de termes, sinó el context. Aquesta capacitat habilita funcions avançades com la desambiguació semàntica (distingint accepcions segons el context), la generació automàtica de resums (utilitzats en determinats àmbits de la descripció de documents) i la normalització de títols.

3.3.1. Aplicació en la catalogació descriptiva i enriquiment

La investigació empírica ha començat a delimitar l'abast real d'aquests models en fluxos de treball bibliotecaris. Un exemple és l'estudi del CRAI de la Universitat de Barcelona (Boté-Vericad, Fabeiro i Anglada, 2024) que ofereix evidència crítica sobre la

implementació professional. La seva anàlisi comparativa revela que, si bé les versions gratuïtes de models com ChatGPT són funcionals per a tasques bàsiques, les versions *premium* són indispensables per a contextos complexos. Mitjançant l'ús d'*Instruccions del Sistema (System Instructions)* precises sobre estàndards RDA i MARC21, aquests models van demostrar capacitat per actuar com a assistents experts, interpretant contextos històrics en documents antics i realitzant transliteracions de llengües complexes (com l'hebreu o el llatí), reduint dràsticament els temps de procés.

Per la seva banda, Brzustowicz (2023) assenyala la capacitat dels LLMs per interpretar dades bibliogràfiques no estructurades i transformar-les en formats estandarditzats. Complementàriament, Jiang et al. (2024) i Husic (2025) destaquen la seva competència en la transliteració i romanització de grafies no llatines (com el xinès o el rus), actuant com a pont eficaç entre dades vernacles i estàndards internacionals.

3.3.2. La «paradoxa de l'eficiència»

Malgrat les capacitats generatives, l'automatització planteja desafiaments qualitatius. Chisaba-Pereira et al. (2025) identifiquen el que es podria denominar una «paradoxa de l'eficiència»: encara que la IA va demostrar ser fins a 183 vegades més ràpida que un catalogador humà, genera registres amb una densitat de dades significativament menor (fent una mitjana de 21 camps MARC enfront dels 55 produïts per un expert). Aquesta troballa suggereix que el rol òptim de la IA actual se situa en fases de pre-catalogació o creació de registres base, requerint intervenció humana per assolir nivells de profunditat descriptiva exhaustius.

3.3.3. Mitigació d'al·lucinacions: arquitectura RAG

Per contrarestar la tendència dels models generatius a l'«al·lucinació» o invenció de dades que no es troben en les fonts originals utilitzades per a la descripció, la indústria avança cap a arquitectures de **Generació Augmentada per Recuperació (Retrieval-Augmented Generation - RAG)**. Cotta García (2025) il·lustra l'eficàcia d'aquest enfocament en el context dels Recursos Educatius Oberts (REO).

El sistema RAG utilitza models d'embeddings per vectoritzar metadades i recuperar informació contextualment similar d'índexs de confiança (com el *Directory of Open Access Journals - DOAJ*) abans de generar una resposta. D'aquesta manera, l'LLM no «inventa» metadades des del buit estadístic, sinó que infereix i completa camps basant-se en dades bibliogràfiques reals recuperades, garantint una descripció fonamentada.

3.3.4. Normalització del text i morfologia

Finalment, és necessari destacar que l'eficàcia del PLN varia segons la tipologia lingüística. Un estudi exhaustiu sobre l'eina *Annif* (Suominen i Koskenniemi, 2022) va demostrar que, per a llengües morfològicament riques com el finès, els mètodes de lematització avançats (que redueixen les paraules a la seva arrel lingüística) superen significativament els algorismes bàsics d'*stemming* (truncament). Això implica que no existeix una solució tecnològica única: la implementació tècnica en biblioteques multilingües requereix una configuració híbrida on l'analitzador lingüístic se seleccioni experimentalment per a cada parell d'idioma i algoritme.

3.4. Reconeixement d'Entitats Nomenades (*Named Entity Recognition - NER*) i control d'autoritats

El Reconeixement d'Entitats Nomenades (*Named Entity Recognition - NER*) constitueix la tecnologia base per a l'automatització del control d'autoritats. En termes tècnics, aquest procés permet analitzar fluxos de text no estructurat (procedents d'OCR o PDF) per identificar, classificar i extreure cadenes de caràcters que corresponen a categories ontològiques redefinides, com persones, institucions, llocs o obres.

La seva aplicació directa en l'entorn MARC21 resideix en la capacitat per incorporar automàticament els punts d'accés dels registres bibliogràfics, discriminant amb precisió si una entitat ha d'assignar-se a un camp d'autoria personal (100, 700), corporativa (710) o a una matèria geogràfica (651).

3.4.1. De la identificació a la reconciliació d'entitats

El valor diferencial del NER en biblioteques no es limita a la mera extracció de text, sinó que habilita el procés de reconciliació d'entitats. Aquest flux transforma cadenes de text ambigües en identificadors inequívocs, connectant el catàleg local amb l'ecosistema global de dades. Existeixen diferents evidències empíriques en aquest sentit.

El flux de treball implementat a la Biblioteca Reial de Bèlgica (KBR) il·lustra l'aplicació industrial d'aquest concepte orientant-lo a l'automatització de l'enriquiment d'autoritats. Segons detalla Lowagie (2023), el sistema extreu mencions de responsabilitat de les cobertes digitalitzades i executa consultes automatitzades (via API HTTP) contra fitxers d'autoritat internacionals com ISNI i VIAF. En detectar una coincidència positiva, el sistema no només transcriu el nom, sinó que enriqueix el registre amb l'Identificador Persistent (URI) corresponent. Aquest mecanisme transforma el catàleg pla en una estructura de *Linked Data*, tasca que manualment resulta inassumible en termes de costos temporals.

D'altra banda, la investigació de Boté-Vericad, Fabeiro i Anglada (2024) amplia l'abast del NER més enllà del text escrit (text i imatge). En els seus experiments amb fons gràfics, van demostrar que els models d'IA multimodal són capaços d'identificar figures històriques i religioses (reis, sants) directament a partir de l'anàlisi visual de gravats antics. Aquesta capacitat permet suggerir encapçalaments de matèria iconogràfics precisos, aportant un nivell d'indexació que complementa la descripció textual tradicional.

3.5. IA simbòlica, vectorització i grafs de coneixement

Enfront de la naturalesa probabilística dels models de llenguatge (que operen realitzant inferències estadístiques), la catalogació requereix la certesa que aporta la **IA Simbòlica**. Aquesta branca de la intel·ligència artificial es fonamenta en l'ús de regles lògiques, ontologies estructurades (com BIBFRAME, RDA-Resource Description and Access o LRM-Library Reference Model) i grafs de coneixement, proporcionant l'estructura necessària per mitigar les "al·lucinacions" inherents als models generatius.

3.5.1. Vectorització (*Embeddings*) i cerca semàntica

La integració d'ambdues aproximacions es realitza mitjançant la **vectorització** (*Embeddings*). Tècnicament, aquest procés converteix el text d'un recurs bibliogràfic en vectors numèrics dins d'un espai multidimensional. A diferència de la cerca per paraules

clau exacta, la vectorització permet identificar «proximitat semàntica». Això habilita sistemes de recuperació que poden localitzar documents conceptualment similars, encara que no comparteixin lèxic idèntic, millorant significativament la capacitat de navegació (i el descobriment) del catàleg.

3.5.2. Arquitectura RAG

Una de les innovacions tècniques més prometedores per a la catalogació és l'arquitectura de **Generació Augmentada per Recuperació** (*Retrieval-Augmented Generation* - RAG). Cotta García (2025) aprofundeix en aquesta metodologia com a solució per a l'enriquiment de metadades en Recursos Educatius Oberts (REO).

El funcionament del sistema RAG es basa a aprofitar la capacitat generativa de l'LLM. En lloc de permetre que el model generi descripcions basant-se exclusivament en les seves dades d'entrenament (que poden estar obsoletes o esbiaixades), el sistema recupera informació d'un índex de confiança extern (com un catàleg d'autoritats, el DOAJ o Wikidata) i l'aplica en el context de la generació. Cotta García (2025) demostra que aquesta tècnica obliga la IA a generar descripcions basades en relacions semàntiques reals, reduint dràsticament la invenció de dades.

3.5.3. Cap a les Dades Enllaçades (*Linked Data*)

Aquesta hibridació tecnològica és el catalitzador per a la transició des de registres MARC plans cap a estructures de Dades Obertes Enllaçades (Linked Open Data). L'objectiu no és només la identificació d'entitats, sinó la seva vinculació a Grafs de Coneixement globals. El projecte de la Biblioteca Reial de Bèlgica (KBR) exemplifica aquesta tendència en generar sortides duals en els seus fluxos d'automatització: produint simultàniament registres en format MARC21xml per als sistemes tradicionals i BIBFRAME-rdfxml per a la web semàntica, assegurant així la interoperabilitat futura.

L'esforç necessari, moltes vegades inassumible amb mitjans humans, per fer realitat l'objectiu de la Web semàntica, té el potencial de reduir-se mitjançant l'ús de tecnologies basades en IA.

3.6. Síntesi de tecnologies i aplicacions

La següent matriu sintetitza la correspondència entre les tècniques d'IA analitzades i la seva aplicació directa en els processos de catalogació:

Tècnica d'IA	Aplicació en catalogació	Exemple d'ús real
OCR + Visió per Computador	Catalogació descriptiva (Camps 245, 260/264)	Escaneig de portada per a extracció automàtica de Títol, Menció d'Edició i Peu d'Impremta.
Reconeixement d'Entitats (NER)	Control d'autoritats (Camps 1XX, 7XX)	Detecció de cadenes de text i enllaç automàtic al seu URI en fitxers d'autoritat (VIAF/ISNI).
Classificació Supervisada (ML)	Anàlisi de matèries (Camps 080, 082)	Algorismes probabilístics (com <i>Annif</i>) que suggereixen signatures.
LLMs (IA Generativa)	Notes i resums (Camps 500, 505, 520)	Generació de resums analítics i notes de contingut a partir del text complet o índexs.
Vector Search (Embeddings)	Recuperació d'Informació	Cerca per similitud semàntica i recomanació contextual.
RAG (Retrieval-Augmented Gen)	Validació i enriquiment	Consulta automatitzada a bases de coneixement (Wikidata) per completar dates o desambiguar autors.

3.7. Prospectiva tecnològica: cap a la IA neuro-simbòlica

La síntesi de les tecnologies analitzades en les seccions anteriors apunta cap a l'aparició d'un nou paradigma arquitectònic: la **IA neuro-simbòlica**. Aquest enfocament no ha d'interpretar-se com una tecnologia independent, sinó com la convergència operativa entre dues branques de la intel·ligència artificial històricament dissociades.

Per una banda, la **IA Neuronal** (representada pels LLMs i la IA Generativa) aporta la capacitat de processament lingüístic, flexibilitat i comprensió del context no estructurat. Per una altra, la **IA Simbòlica** aporta la fiabilitat, la lògica basada en regles i l'estructura jeràrquica dels grafs de coneixement i les ontologies bibliotecàries (com BIBFRAME o RDA).

La literatura tècnica suggereix que el futur de l'automatització bibliotecària resideix en la integració d'ambdós models. Aquesta convergència permet desenvolupar sistemes capaços de «llegir» i «escriure» amb la fluïdesa d'un model generatiu, però obligats a «verificar» les seves inferències contra estructures de dades controlades. Aquesta arquitectura híbrida és la resposta tècnica definitiva per mitigar el risc d'al·lucinació i corrupció de dades, garantint que l'eficiència de l'automatització no comprometí la integritat i coherència del catàleg.

4. Context institucional: adopció i marcs de referència a Europa

Per comprendre l'estat de la qüestió més enllà de l'oferta comercial, és necessari analitzar la realitat operativa de les institucions. Aquesta anàlisi es fonamenta en les dades de les associacions professionals que vertebren la cooperació bibliotecària al continent (CENL, LIBER) i en el context espanyol (REBIUN).

4.1. L'escenari europeu: entre l'experimentació i la producció

A nivell continental, les dades de la Conference of European National Librarians (CENL) matisen la narrativa d'una adopció tecnològica massiva. Si bé l'informe global Pulse of the Library 2025 (Clarivate, 2025) suggereix que el 67% de les biblioteques a nivell mundial estan explorant la IA, l'enquesta específica de la CENL (2024) revela una bretxa significativa en la implementació real.

Encara que el 88,5% de les biblioteques nacionals europees reporta activitat relacionada amb la Intel·ligència Artificial, només un 29% disposa de sistemes en fase de producció. Aquesta dada evidencia que Europa es troba en una fase de transició: existeix una avantguarda de «sobirania tecnològica» liderada per institucions del nord (com la Biblioteca Nacional de Finlàndia i el seu projecte Annif, o la de Noruega), mentre que gran part del continent continua avaluant la viabilitat tècnica i ètica d'aquestes eines abans d'integrar-les en els seus catàlegs oficials.

4.2. L'Observatori d'Intel·ligència Artificial de REBIUN

En el context espanyol, la referència fonamental és l'Observatori d'Intel·ligència Artificial de REBIUN. Constituït amb objectius operatius el 2024, aquest organisme ha actuat com un regulador d'expectatives, promovent una anàlisi crítica sobre l'aplicabilitat de la IA.

L'informe anual Intel·ligència artificial en biblioteques universitàries espanyoles (REBIUN, 2024) descriu un ecosistema en fase incipient. L'adopció d'IA en processos tècnics de catalogació se situa al voltant del 6%, predominant l'experimentació sobre la implementació productiva. Tal com va exposar Franganillo (2024) en el marc de l'assemblea anual, el sector espanyol ha transitat d'una «**eufòria tecnològica**» inicial cap a una «**prudència estratègica**». Les prioritats actuals, alineades amb la missió de l'Observatori, se centren en la formació del personal i l'establiment de directrius ètiques i legals abans d'escometre l'automatització massiva de la descripció bibliogràfica.

5. Anàlisi del sector tecnològic i aplicació en processos tècnics

Una vegada establert el context institucional, aquest apartat analitza l'oferta del mercat privat i, posteriorment, desglossa com aquestes tecnologies s'estan aplicant a processos bibliotecaris específics.

5.1. El mercat global de tecnologia bibliotecària

L'anàlisi del sector privat es basa en la sèrie de treballs titulats *Library Systems Report* (2023-2025) dirigida per Marshall Breeding. L'examen d'aquests informes revela un mercat caracteritzat per una consolidació corporativa extrema i una estratègia de desplegament tecnològic molt específica.

5.1.1. L'estratègia de la «IA Invisible»

En el seu informe de 2024, Breeding declara l'inici de l'«Era de la IA» a les biblioteques, però aporta un matís crucial: la Intel·ligència Artificial no es comercialitza majoritàriament com un producte independent, sinó que s'integra de manera «invisible» a les Plataformes de Serveis Bibliotecaris (LSP) existents. Les institucions no solen adquirir programari específic d'«IA per catalogar», sinó subscripcions a plataformes consolidades (com *Alma* d'Ex Libris/Clarivate o *WorldShare* d'OCLC) que han començat a incorporar algorismes d'aprenentatge automàtic (*Machine Learning*) en els seus fluxos de treball nadius per a tasques de deduplicació i enriquiment de metadades (Breeding, 2024).

5.1.2. Consolidació i divergència tecnològica

El mercat s'estructura en un oligopoli dominat per gegants com Clarivate i OCLC, observant-se una bifurcació tècnica clara (Breeding, 2025):

- Sector acadèmic. Ha completat la seva migració a plataformes LSP (*Library Services Platforms*) on la IA té un alt impacte potencial a causa de l'escala massiva de dades compartides.
- Resta del sector públic. Manté majoritàriament ILS (*Integrated Library Systems*) tradicionals, on la innovació en IA se centra més en la interfície de descobriment que en el processament tècnic (*back-end*).

5.2. Catalogació automatitzada: de la ingesta al registre estructurat

L'horitzó operatiu de l'automatització bibliotecària consisteix en la capacitat del sistema per generar un registre bibliogràfic complet directament a partir de l'anàlisi de l'objecte (físic o digital) amb una intervenció humana mínima. La literatura tècnica recent desglossa aquest procés en tres fluxos de treball diferenciats: la ingesta d'arxius nadius digitals, la catalogació visual de suports físics i la identificació de manifestacions.

5.2.1. Ingesta de documents digitals (PDF a MARC)

Per als recursos nascuts digitals, l'aplicació de Grans Models de Llenguatge (LLMs) ha demostrat ser eficaç en l'extracció estructural. Husic (2025) documenta com els models generatius poden processar arxius PDF o EPUB, analitzant la seva morfologia interna (portada, pàgina de *copys*, taula de continguts, etc.) per mapejar les dades extretes a les etiquetes MARC corresponents (245, 100, 264, etc.). Aquesta capacitat resulta crítica

per a la gestió de literatura grisa, tesis i recursos educatius oberts que manquen de catalogació a la font.

5.2.2. Catalogació Visual i fluxos híbrids

En el tractament de fons físics, la combinació de Visió per Computador i automatització de processos robòtics (RPA) ha permès industrialitzar la creació de registres:

- **Fluxos Industrials (KBR).** La Biblioteca Reial de Bèlgica ha establert un paradigma operatiu mitjançant l'ús de *Microsoft Power Automate*. Segons descriu Lowagie (2023), el flux s'activa quan el personal escaneja el llibre i diposita la imatge en una carpeta al núvol. El sistema processa visualment l'arxiu, extreu les dades mitjançant models OCR entrenats i retorna un registre esborrany al Sistema Integrat de Gestió Bibliotecària (SIGB), reduint el temps de processament tècnic de manera substancial.
- **Fons Patrimonials (CRAI-UB).** En contextos d'alta complexitat, com el fons antic, Boté-Vericad, Fabeiro i Anglada (2024) evidencien que els models avançats (GPT-4) poden transcriure títols amb grafies arcaïques i identificar mencions de responsabilitat en portades barroques, proposant una estructura MARC21 que inclou camps de notes (500) inferits de l'anàlisi visual.

5.2.3. Identificació de la manifestació i gestió d'adquisicions

La IA no només opera en la creació, sinó en la discriminació del recurs. Ng i Goh (2025), de la National Library Board de Singapur, presenten l'eina SGCAT, dissenyada per processar llistes d'adquisicions. L'algorisme verifica l'existència del recurs contra bases de dades globals (com WorldCat) i, exclusivament davant l'absència de coincidències, genera un registre bibliogràfic preliminar ric en metadades. Aquest enfocament transforma el rol del catalogador, que passa de la creació *ex novo* a la validació, reduint el temps de catalogació.

5.2.4. Limitacions empíriques i dependència normativa

Malgrat l'eficiència reportada, la validació experimental alerta sobre deficiències qualitatives. L'estudi de Selbach et al. (2024) a la Pontifícia Universidade Católica do Rio Grande do Sul va revelar que, sense un entrenament específic, els models generalistes (com ChatGPT 3.5) tendeixen a prioritzar regles angloamericanes (AACR2) sobre la normativa RDA vigent. Així mateix, aquests models van mostrar una propensió a l'«al·lucinació» en tasques de classificació decimal, inventant notacions inexistents.

Paral·lelament, Gamage i Wanigasooriya (2024) assenyalen barreres en el reconeixement òptic de caràcters (OCR) per a llengües amb menors recursos digitals, com el singalès o el tàmil, on la IA falla en el processament de les imatges de coberta, limitant l'aplicabilitat global d'aquestes eines sense un desenvolupament local específic.

5.3. Implementació assistida d'RDA i model LRM

La transició des de les regles angloamericanes tradicionals cap a la normativa RDA (Resource Description and Access) ha suposat un increment en la complexitat cognitiva del procés de catalogació, en transitar d'una descripció plana a un model basat en entitats i relacions (LRM). En aquest domini, la Intel·ligència Artificial es perfila com una eina eficaç per a la gestió de l'estructura ontològica subjacent.

Inferència de relacions WEMI La capacitat de la IA semàntica resulta particularment apta per operacionalitzar el model conceptual WEMI (Obra, Expressió, Manifestació, Ítem). A diferència dels sistemes basats en regles rígides, els models de llenguatge poden inferir relacions complexes a partir de l'anàlisi del text. Per exemple, davant d'un recurs identificat com a «Don Quijote (trad. anglès)», un model adequadament entrenat pot suggerir la seva vinculació automàtica com una Expressió de l'Obra original cervantina, assignant els designadors de relació RDA pertinents (per exemple, traduït per) sense intervenció manual específica.

Automatització d'atributs tècnics i puntuació ISBD La investigació empírica valida la utilitat dels assistents virtuals per reduir la corba d'aprenentatge tècnica d'RDA. L'estudi realitzat pel CRAI de la Universitat de Barcelona (Boté-Vericad, Fabeiro i Anglada, 2024) va demostrar que, mitjançant l'ús d'instruccions del sistema (*Custom Instructions*) personalitzades amb les polítiques de catalogació de la institució, la IA és capaç de generar amb alta precisió els camps codificats 336, 337 i 338 (tipus de contingut, suport i suport físic) basant-se únicament en l'anàlisi físic o visual de l'objecte.

Així mateix, aquest estudi destaca la competència de la IA en l'estructuració de la menció de responsabilitat, respectant rigorosament la puntuació ISBD prescrita. Aquesta capacitat automatitzada per gestionar la sintaxi formal allibera el catalogador de tasques mecàniques de validació de format, permetent-li concentrar-se en la identificació intel·lectual de les entitats.

5.4. Anàlisi de contingut: indexació, classificació i resum

L'anàlisi de contingut representa la dimensió «intel·lectual» de la catalogació, tradicionalment dependent de la interpretació humana. En aquest àmbit, la tecnologia ha evolucionat des de l'anàlisi de freqüència de paraules cap a sistemes híbrids que combinen algorismes lèxics, aprenentatge automàtic i models generatius.

5.4.1. Indexació semàntica automatitzada: l'estàndard Annif

L'eina de codi obert *Annif*, desenvolupada per la Biblioteca Nacional de Finlàndia, s'ha consolidat com la solució de referència per a la indexació automàtica de matèries. La seva arquitectura permet la implementació d'«assemblatges» (*ensembles*), combinant algorismes associatius i lèxics per equilibrar la precisió terminològica amb la recuperació estadística.

La participació recent en l'avaluació competitiva GermEval-2025 ha evidenciat la superioritat dels enfocaments híbrids. Suominen, Inkinen i Lehtinen (2025) han demostrat que l'estratègia òptima no consisteix a utilitzar LLMs per a la predicció directa de matèries, sinó com a agents auxiliars per a la traducció automàtica i la generació de dades d'entrenament sintètiques. Aquesta tècnica permet enriquir conjunts de dades escassos, millorant el rendiment d'algorismes de classificació tradicionals.

5.4.2. Classificació automàtica: precisió i limitacions disciplinàries

L'assignació de signatures topogràfiques (basades en CDU, Dewey...) mitjançant aprenentatge supervisat mostra resultats variables segons el domini del coneixement. Chadwick (2022), en el seu treball amb *Schools Catalogue Information Service (SCIS)*, va validar que classificadors estadístics poden alinear recursos amb àrees curriculars amb una precisió superior al 90% en categories ben representades.

No obstant això, la complexitat augmenta en col·leccions generalistes. Un estudi sobre el Catàleg Col·lectiu Suec (*Libris*), que va analitzar més de 230.000 registres, va reportar una precisió del 66,82% en l'assignació de codis Dewey a nivell de tres dígits utilitzant *Annif*. Golub et al. (2024) destaquen una troballa crítica: els algorismes presenten una taxa d'error sistemàtica en obres de ficció, confonent freqüentment l'escenari geogràfic o cultural de la novel·la amb la seva matèria, cosa que suggereix que les metadades tradicionals són insuficients per entrenar models en literatura imaginativa sense l'anàlisi del text complet.

5.4.3. Generació de resums i enriquiment contextual

On la IA Generativa mostra una competència distintiva és en la redacció de resums i notes d'abast. Cotta García (2025) il·lustra com l'ús d'embeddings permet generar resums coherents per a recursos que manquen d'ells, facilitant la seva recuperació semàntica.

Paral·lelament, l'estudi del CRAI-UB (Boté-Vericad, Fabeiro i Anglada, 2024) va revelar la capacitat de la IA per aportar valor expert en fons patrimonials. En les seves proves, els models van ser capaços de generar descripcions detallades de gravats històrics, identificant iconografia religiosa i contextos específics que un catalogador generalista podria ometre sense una investigació extensa.

5.5. Control de Qualitat i Manteniment del Catàleg (*Data Hygiene*)

La integritat de les metadades constitueix un requisit funcional de vital importància per a la recuperació de la informació i per a la coherència i qualitat dels catàlegs. En aquest domini, la Intel·ligència Artificial actua com un mecanisme d'auditoria massiva, superant les limitacions dels algorismes tradicionals en tasques de depuració i normalització.

5.5.1. Detecció de duplicats difusos

A diferència dels Sistemes Integrats de Gestió Bibliotecària (SIGB) convencionals, que requereixen coincidències exactes de caràcters per identificar duplicitats, els models d'aprenentatge automàtic empenen lògica difusa i *clustering semàntic*. Aquesta capacitat permet al sistema detectar que variants d'un nom d'un mateix autor corresponen a la mateixa identitat, facilitant la fusió de registres duplicats fins i tot davant la presència d'errors tipogràfics o variacions en la puntuació.

Grans proveïdors d'infraestructura bibliotecària, com OCLC, estan implementant aquestes tècniques a escala global a WorldCat per sanejar la base de dades col·lectiva, identificant manifestacions idèntiques que han estat catalogades de manera dispar per diferents institucions.

5.5.2. Normalització i neteja massiva

Per a la correcció retrospectiva, la integració de Models de Llenguatge (LLMs) en eines de manipulació de dades com OpenRefine ha democratitzat l'accés a la neteja semàntica. Mitjançant extensions d'API, és possible processar columnes de dades «brutes» sol·licitant a la IA la seva normalització cap a una forma autoritzada o expandida.

5.5.3. Validació post-generació

No obstant això, l'automatització del control de qualitat no eximeix de la supervisió tècnica. Ng i Goh, en la seva anàlisi de l'eina SGCAT, destaquen la necessitat d'implementar protocols de validació automatitzada post-generació. En el seu flux de treball, els registres generats per IA són sotmesos a *scripts* de verificació per assegurar la integritat estructural (presència de camps obligatoris, codificació correcta d'indicadors) abans de la seva ingesta definitiva al catàleg.

5.6. Interoperabilitat i Web semàntica: la transició cap a Linked Data

La irrupció de la intel·ligència artificial ha reactivat el debat sobre la vigència dels estàndards bibliotecaris tradicionals en l'era del *Big Data*. La controvèrsia sorgeix en el qüestionament de la necessitat d'invertir recursos en la construcció i manteniment d'ontologies complexes com BIBFRAME o RDA (i en l'adaptació dels registres descriptius a aquestes ontologies) donat el potencial d'inferència dels algorismes moderns, planteja una disjuntiva estratègica per a les biblioteques.

5.6.1. El debat al voltant de les ontologies bibliotecàries: és necessària l'ontologia en l'era dels LLMs?

Certs corrents argumenten que els models de llenguatge actuals poden extreure significat i relacions directament de textos no estructurats amb suficient precisió probabilística, suggerint l'obsolescència d'estructures rígides. No obstant això, aquesta postura ha estat analitzada per Martínez-Ávila (2023) que sosté que renunciar a l'estructura ontològica explícita per confiar únicament en inferències estadístiques constitueix un error estratègic.

Sota la premissa que «el contrari d'ontologia no és no-ontologia, sinó mala ontologia», Martínez-Ávila (2023) argumenta que la transició a BIBFRAME no és merament un canvi de format, sinó la construcció de la infraestructura semàntica necessària perquè la IA pugui raonar de manera fiable. Lluny de fer innecessaris els estàndards, els models intel·ligents requereixen dades d'entrenament d'alta qualitat, estructurades i desambiguades, precisament el que ofereixen els entorns de *Linked Open Data* per operar de manera ètica i mitigar al·lucinacions.

5.6.2. Conversió intel·ligent: de cadenes de text a URI

En el pla tècnic, l'aplicació de la IA està superant les limitacions dels conversors de regles tradicionals (XSLT). La conversió de registres MARC a BIBFRAME no es limita a un mapeig d'etiquetes, sinó que requereix una anàlisi semàntica per transformar cadenes de caràcters en entitats identificades unívocament mitjançant URI.

Breeding (2025) destaca que aquesta capacitat d'enriquiment és una de les àrees de major inversió per part de proveïdors de plataformes (LSP) com EBSCO i OCLC. L'objectiu és automatitzar la creació de grafs de coneixement on les entitats (autors, obres, matèries) estiguin interconnectades, permetent una navegació semàntica que transcendeix la recuperació per coincidència de text.

6. Desafiaments ètics i problemes tecnològics: al·lucinacions, biaixos i marc legal

Malgrat el potencial de l'automatització mitjançant Intel·ligència Artificial, la literatura acadèmica recent convergeix en una postura de cautela. La implementació de la Intel·ligència Artificial en els processos de descripció bibliogràfica no està exempta de riscos estructurals que amenacen la integritat dels catàlegs. L'anàlisi d'aquestes barreres s'articula al voltant de tres eixos crítics: la fiabilitat de les dades (al·lucinacions), l'ètica de la representació (biaixos) i la seguretat jurídica.

6.1. La problemàtica de les al·lucinacions

El risc més documentat en l'ús de Models de Llenguatge (LLMs) resideix en la seva naturalesa probabilística: són màquines dissenyades per maximitzar la versemblança lingüística, no la veracitat factual. Aquest fenomen, conegut com a «al·lucinació», implica la generació de dades falses presentades amb alta confiança estadística.

Engel et al. (2025) descriuen les al·lucinacions com un problema en la generació de metadades, on la IA fabrica fets plausibles però inexistents. L'evidència empírica és abundant:

- **Invenció d'encapçalaments.** Jiang et al. (2024) van documentar casos on ChatGPT va inventar encapçalaments de matèria de la *Library of Congress* (LCSH) lingüísticament coherents però absents en el vocabulari controlat.
- **Errors estructurals.** Encara que el registre sembli correcte a simple vista, la inspecció tècnica revela fallades en camps crítics. Husic (2025) identifica la fragilitat del camp de longitud fixa 008 i la capçalera, on la IA falla sistemàticament en ubicar els espais en blanc i codis posicionals requerits pels sistemes bibliotecaris, impeding la importació vàlida del registre.
- **Incoherència en subdivisions.** Així mateix, els models mostren dificultats per distingir la funció semàntica d'un terme geogràfic, confonent quan ha de tractar-se com a tema principal (camp 651) o com a subdivisió (camp 650 \$z).

6.2. Biaixos heretats i limitacions culturals

La qualitat de la catalogació automatitzada és directament proporcional a la qualitat de les dades d'entrenament (*Ground Truth*). Atès que els models aprenen de catàlegs històrics, existeix un risc elevat de perpetuar biaixos.

- **Biaix cultural i terminològic.** Alemu (2025) alerta sobre el biaix heretat: els models reproduïen terminologia obsoleta, colonial o discriminatòria present en registres antics si no s'apliquen filtres ètics rigorosos.
- **Biaix geogràfic.** En termes d'estàndards, Gamage i Wanigasooriya (2023) van observar una preferència algorítmica per les pràctiques de catalogació occidentals. En els seus experiments, la IA va prioritzar etiquetes MARC utilitzades predominantment en biblioteques anglosaxones en detriment d'esquemes locals més rellevants per al context de la publicació.

6.3. La paradoxa de l'eficiència i el cost operatiu

La promesa de reducció de costos es veu qüestionada pel que Chisaba-Pereira et al. (2025) denominen la «paradoxa de l'eficiència». Si bé la IA és capaç de generar registres a una velocitat 183 vegades superior a la humana, la densitat informativa d'aquests registres és significativament menor (fent una mitjana de 21 camps enfront dels 55 d'un expert).

Aquesta disparitat obliga a replantejar el flux de treball. Miller-Nesbitt (2024) argumenta que el temps requerit pels catalogadors humans per revisar, corregir al·lucinacions i completar els camps que falten en registres generats per IA pot anul·lar els estalvis de temps inicials, fent que l'automatització sigui ineficient per a materials que requereixen una descripció profunda.

6.4. Incertesa jurídica i propietat intel·lectual

Finalment, la integració de la IA planteja interrogants legals encara no resolts per la legislació europea. Brzustowicz (2023) assenyala l'ambigüitat sobre la propietat intel·lectual dels registres catalogràfics generats per màquines i el risc d'infracció de drets d'autor en processar descripcions detallades d'obres protegides.

A nivell d'infraestructura, Storti (2024) destaca una barrera estructural crítica en el context europeu: la falta de normatives operatives de dipòsit legal digital. Sense un marc que obligui i reguli el dipòsit de recursos digitals nadius, les biblioteques manquen de la legitimitat legal i el volum de dades necessaris per entrenar models propis (*Ground Truth*) en llengües locals, forçant una dependència de models comercials entrenats amb dades genèriques i opaques.

7. Estudi de camp: enquesta a institucions europees sobre l'ús de la intel·ligència artificial en projectes de catalogació de documents

7.1. Perfil dels enquestats

L'estudi de camp va ser respost per 87 institucions bibliotecàries de diferents tipologies i grandàries, oferint una visió preliminar del panorama europeu en l'ús de la intel·ligència artificial (IA) en projectes de catalogació de documents. L'anàlisi del perfil dels participants és crucial per contextualitzar les conclusions subsegüents sobre els nivells d'adopció, les àrees d'aplicació i les barreres detectades.

La participació en l'enquesta s'estén a un total de nou països europeus. No obstant això, s'observa una marcada concentració de respostes a Espanya, que aporta 77 de les 87 institucions, la qual cosa representa un aclaparador 88.5% del total de la mostra.

Els altres països contribueixen amb un nombre significativament menor de respostes: Eslovènia (2), Itàlia (2), i amb una única institució cadascun, Bulgària, Eslovàquia, Finlàndia, Geòrgia i Letònia.

Aquesta distribució geogràfica és un factor fonamental que s'haurà de tenir en compte en la interpretació dels resultats, ja que la forta presència espanyola podria esbiaixar les conclusions cap a les particularitats i el context del sector bibliotecari d'aquest país.

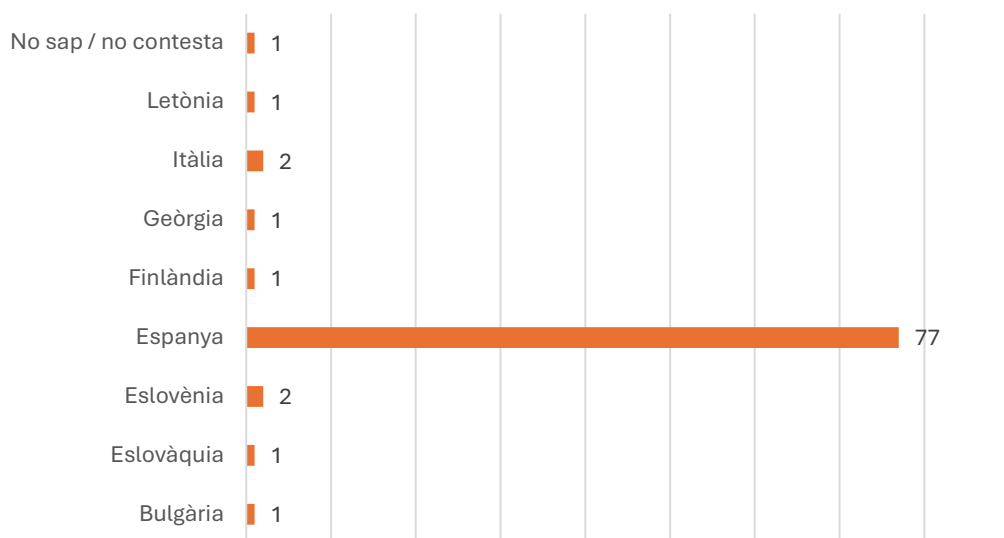


Gráfico 1. Distribució de països participants en l'enquesta

Quant a la naturalesa de les institucions enquestades, s'observa que la tipologia més representada és la Pública, amb 43 respostes. Aquest grup conforma gairebé la meitat de la mostra i és seguit per les biblioteques Especialitzades (19), les Universitàries / d'Investigació (13), i amb una presència idèntica, les biblioteques Locals / Municipals i Nacionals (10 cadascuna). Les tipologies Privada (7), Regional (2) i Altra (especificar) (2) completen la distribució.

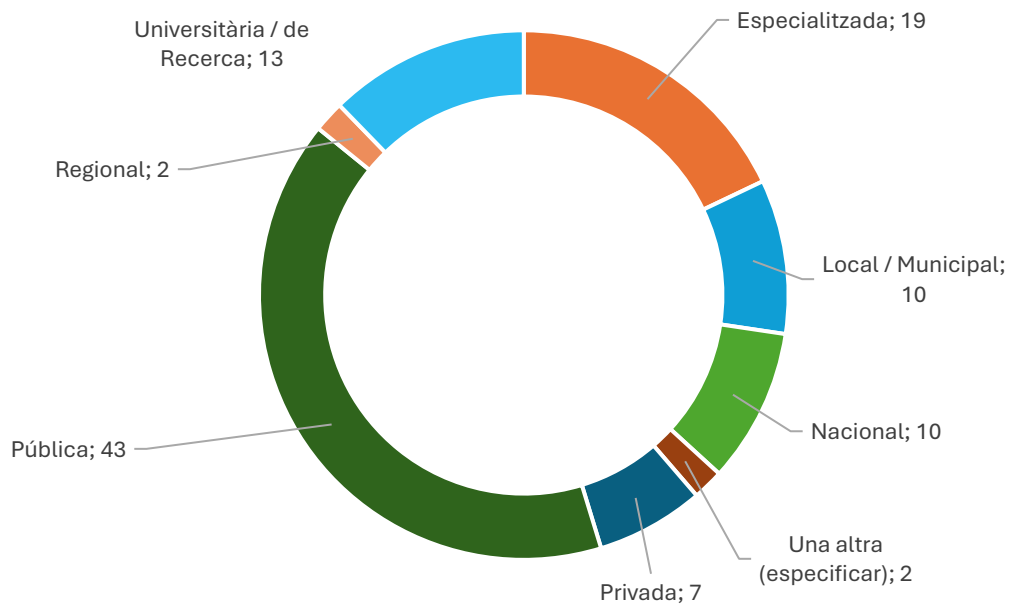


Gráfico 2. Distribució de participants per tipologia de biblioteca

L'anàlisi de la grandària de les col·leccions de les institucions participants reforça el perfil de la mostra. La immensa majoria de les institucions (46) reporta tenir una col·lecció de Menys de 50.000 unitats/ítems, la qual cosa suggereix una alta proporció de biblioteques petites o especialitzades, en línia amb l'alta participació de biblioteques públiques i especialitzades.

La segueixen, a una distància considerable, aquelles amb col·leccions d'entre 50.001 i 250.000 unitats (17), i el rang d'1.000.001 - 5.000.000 (9). Les grans col·leccions, aquelles amb Més de 5.000.000 d'unitats, són el grup més petit, amb 5 institucions. Dos participants no van proporcionar dades sobre la grandària de la seva col·lecció.

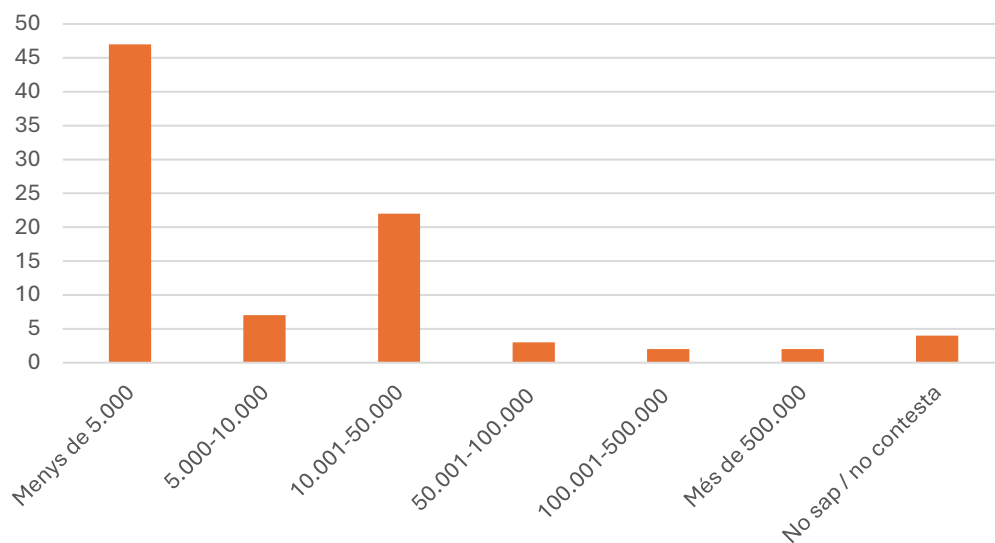
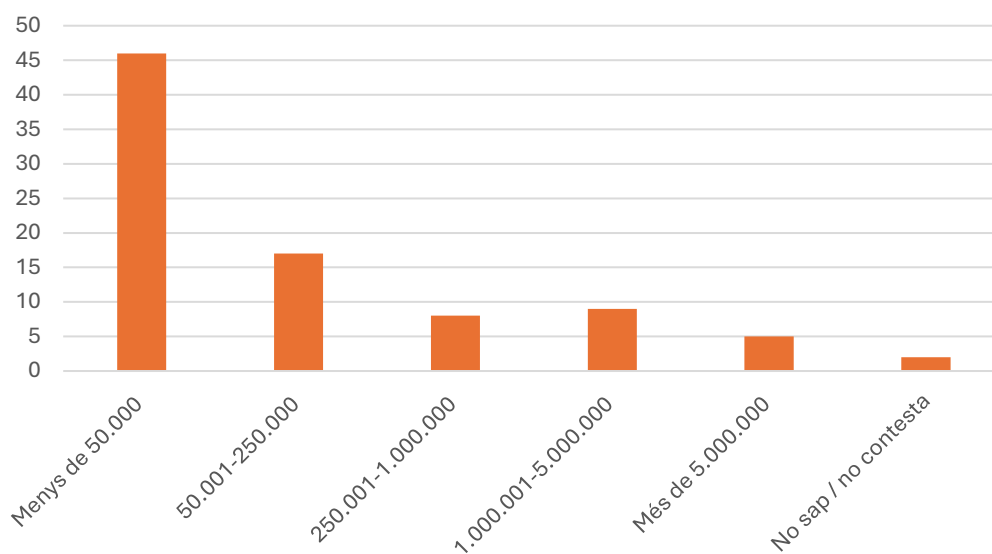


Gráfico 3. Distribució per mesura aproximada de la col·lecció

En consonància amb la grandària de la col·lecció, la majoria de les institucions (47) declaren tenir Menys de 5.000 usuaris registrats o actius. Aquesta dada consolida la tesi que la mostra està dominada per entitats amb un abast geogràfic o temàtic més acotat.

El segon grup més gran (22) se situa en el rang de 10.001 - 50.000 usuaris, la qual cosa pot correspondre a biblioteques universitàries o municipals de grandària mitjana. Els rangs d'usuaris més alts (més de 100.000) tenen una representació mínima (2 en cadascun), la qual cosa suggereix que les grans biblioteques nacionals o amb un gran nombre d'usuaris actius han participat en menor mesura. Quatre participants no van contestar a aquesta pregunta.



Gràfic 4. Distribució per rang aproximat d'usuaris

Per tant, el perfil de la institució mitjana enquestada és el d'una biblioteca pública o especialitzada d'Espanya, amb una grandària de col·lecció petita (menys de 50.000 unitats) i un nombre d'usuaris reduït (menys de 5.000). Aquesta sobre-representació d'un tipus específic d'institució i país suggereix que els resultats sobre l'ús de la IA en catalogació podrien reflectir de manera més fidel les prioritats i desafiaments d'aquest segment del sector, i menys el de les grans biblioteques nacionals o d'investigació a nivell europeu.

7.2. Nivell d'adopció actual i àrees d'aplicació prioritària

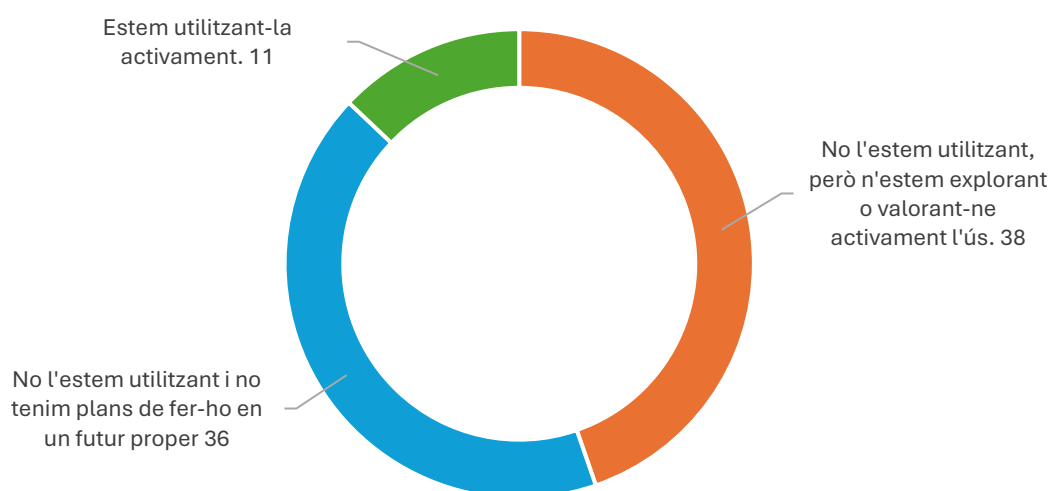
L'anàlisi de les dades de l'enquesta revela que la Intel·ligència Artificial (IA) en les institucions bibliotecàries europees es troba majoritàriament en una fase d'exploració i valoració, amb una minoria que reporta un ús actiu i una porció significativa que encara no contempla la seva implementació.

Les biblioteques europees mostren un interès significatiu i un nivell d'adopció incipient però tangible de la Intel·ligència Artificial, principalment enfocat en tasques que faciliten l'atenció a l'usuari, la gestió de contingut i la catalogació avançada. L'adopció es troba en una fase on l'exploració i la valoració de futur ús tenen un pes considerable, superant en molts casos l'ús actual ja implementat.

7.2.1. Adopció actual de la IA

La distribució de les respostes sobre la situació actual en l'ús de la IA mostra una clara polarització:

- No l'estem utilitzant, però estem explorant o valorant activament el seu ús: Aquesta és la categoria dominant, amb 38 mencions. Això indica un alt interès potencial i una etapa d'estudi de viabilitat per part de la majoria de les institucions.
- No l'estem utilitzant i no tenim plans de fer-ho en un futur pròxim: Un nombre considerable d'institucions, 36, es troben en aquesta situació. Això suggereix que, malgrat l'interès general en el sector, una part important de la mostra encara no ha integrat la IA en el seu full de ruta estratègic.
- L'estem utilitzant activament: Només 11 institucions reporten estar utilitzant activament la IA en els seus fluxos de treball. Encara que aquesta xifra és la més baixa, representa una base d'usuaris implementada l'experiència dels quals pot ser crucial per a l'adopció futura.



Gràfic 5. Situació actual respecte a l'ús de la IA

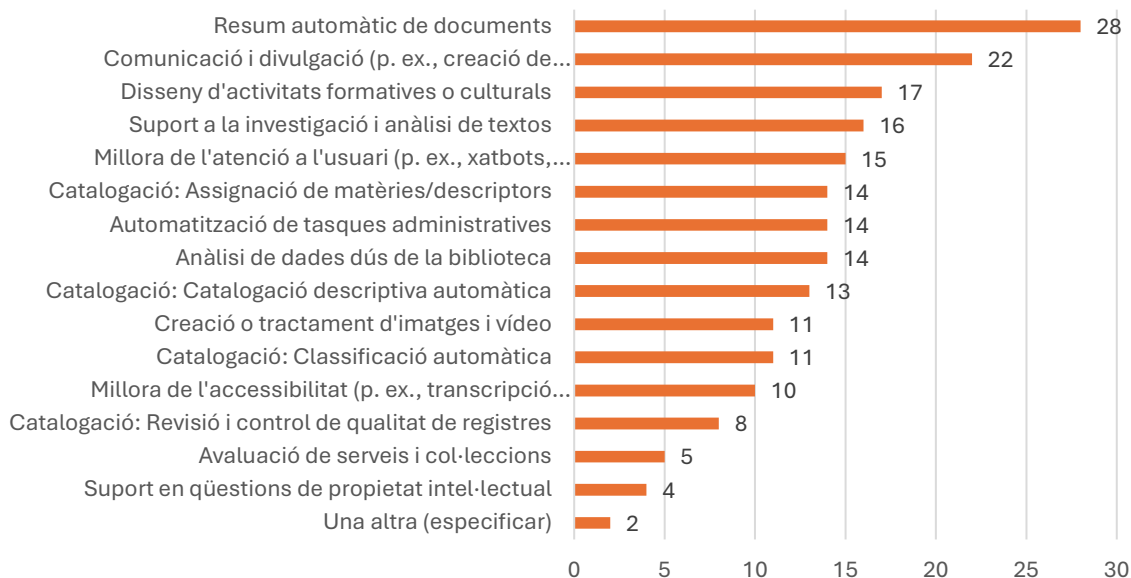
En consonància amb la fase d'exploració predominant, la fase d'aplicació en el flux de treball intern de les institucions que utilitzen o exploren la IA es reparteix de la següent manera:

- No s'aplica: 20 institucions.
- Fase d'exploració / consideració teòrica: 18 institucions.
- Projecte pilot / Fase de proves: 9 institucions.
- En producció (integrat en el treball diari): Només 2 institucions.

L'escassa quantitat d'institucions en la fase d'«En producció» (2) subratlla que la IA és encara una tecnologia incipient o experimental en el sector bibliotecari europeu, amb la majoria dels esforços centrats en la investigació i el pilotatge.

7.2.2. Ús i valoració de la IA: àmbits d'interès

L'interès de les institucions enquestades es polaritza cap a àrees que prometen un alt impacte en l'eficiència operativa i en la millora dels serveis. Les dades sobre l'Ús o valoració d'ús de la IA identifiquen tres dominis principals:



Gràfic 6. Ús o valoració d'ús de la IA

- **Gestió de Contingut i Documents:** Les tasques relacionades amb el resum automàtic de documents (28 mencions) i el suport a la investigació i anàlisi de textos (16 mencions) lideren l'interès. Això suggereix una clara necessitat d'eines que ajudin a processar grans volums d'informació i a facilitar l'accés temàtic al contingut.
- **Atenció a l'Usuari i Serveis:** La millora de l'atenció a l'usuari (p. ex., chatbots, sistemes de recomanació) amb 15 mencions, i el disseny d'activitats formatives o culturals amb 17 mencions, demostren que la IA es percep com un motor per enriquir la interacció amb el públic i la programació de la biblioteca.
- **Catalogació i Tasques Administratives:** L'interès a automatitzar processos interns és notable, amb la catalogació (assignació de matèries/descriptius, catalogació descriptiva automàtica, classificació automàtica) acumulant 38 mencions en total. Així mateix, l'automatització de tasques administratives (14 mencions) i l'anàlisi de dades d'ús de la biblioteca (14 mencions) evidencien un desig d'optimitzar la gestió interna.

7.2.3. Tècniques d'IA explorades o implementades

Respecte a les tecnologies específiques, les biblioteques se centren en aquelles que ja han demostrat maduresa i utilitat en el processament de text i l'automatització. Les dades sobre Tècniques d'IA usades o valorades per posar en ús són clares:

Tècniques d'IA usades o valorades per posar en ús	
Traducció automàtica (p. ex., per traduir metadades o documents)	29
Reconeixement Òptic de Caràcters (OCR) (p. ex., per digitalitzar i fer cercable el text de documents impresos)	25
Models de Llenguatge Grans (LLM) i chatbots (p. ex., per respondre preguntes d'usuaris o generar text)	25
Generació d'imatges (IA generativa) (p. ex., per crear il·lustracions per a esdeveniments o publicacions)	22
Automatització Robòtica de Processos (RPA) (p. ex., per automatitzar tasques repetitives de programari)	20
Processament del Llenguatge Natural (PLN): tècniques que permeten a les màquines comprendre i processar el llenguatge humà	17
Sistemes de recomanació (p. ex., per suggerir llibres o altres recursos als usuaris)	15
Reconeixement d'objectes i imatges (per exemple, per identificar i etiquetar automàticament elements en fotografies o vídeos)	14
Reconeixement d'entitats nomenades (NER) (per exemple, per identificar automàticament autors, llocs o temes en un text)	12
Aprenentatge automàtic (Machine Learning) i anàlisi predictiva: algorismes que aprenen de les dades per fer prediccions o prendre decisions	6
Reconeixement de veu (Speech-to-Text) (per exemple, per transcriure fitxers d'àudio)	5
Anàlisi predictiva de la demanda (p. ex., per anticipar quins materials seran més sol·licitats)	4
Visió per ordinador: tècniques que permeten a les màquines interpretar i comprendre informació visual del món	3
Detecció d'anomalies (per exemple, per identificar plagi o usos indeguts de recursos electrònics)	2
Automatització intel·ligent i robòtica	2
Anàlisi de sentiments (p. ex., per avaluar opinions d'usuaris sobre serveis o col·leccions)	2
Robots físics (p. ex., per a la gestió de magatzems o l'atenció al públic)	1

- Processament de Llenguatge i Visió: El major interès es concentra en la Traducció automàtica (29 mencions) i el Reconeixement Òptic de Caràcters (OCR) (25 mencions). Aquestes tècniques són fonamentals per a projectes de digitalització i per facilitar l'accés multilingüe a les col·leccions.
- IA Generativa i Interacció: Els Models de Llenguatge Grans (LLM) i chatbots obtenen 25 mencions, igualant a l'OCR, la qual cosa subratlla la importància de la IA generativa no només per a l'atenció a l'usuari, sinó també per a tasques de generació de text i metadades. La Generació d'imatges (IA generativa) també té una alta valoració amb 22 mencions.
- Automatització i Anàlisi: Tècniques com l'Automatització Robòtica de Processos (RPA) (20 mencions) i el Processament del Llenguatge Natural (PLN) (17 mencions) són valorades per optimitzar fluxos de treball repetitius i per a la comprensió semàntica dels continguts.

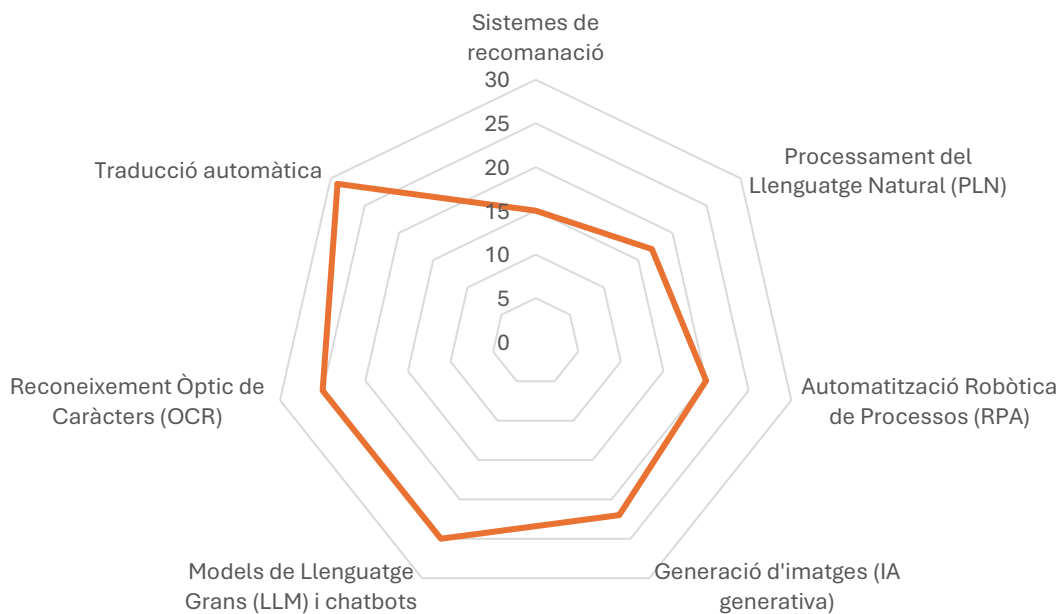


Gráfico 7. Tècniques d'IA més utilitzades o explorades

7.2.4. Tasques específiques d'implementació de la IA actualment

En examinar les tasques en les quals s'està utilitzant la IA, es confirma la concentració d'esforços en el nucli de la gestió de col·leccions i l'accés a la informació:

- **Catalogació i Classificació:** La creació de registres bibliogràfics complets des de zero (catalogació original) i l'assignació de notacions de classificació (CDU, Dewey, LCC, etc.) són les tasques més citades, ambdues amb 23 mencions. Aquestes xifres desmenteixen parcialment la idea d'un baix ús de la IA en catalogació, evidenciant que aquesta àrea és, de fet, el principal focus d'aplicació actual.
- **Indexació i Accés Temàtic:** La segueixen de prop l'assignació d'encapçalaments de matèria, descriptors o paraules clau (20 mencions) i la generació de resums i notes de contingut (18 mencions). La IA s'utilitza, per tant, per enriquir les metadades i millorar la indexació.
- **Control de Qualitat:** La revisió, correcció i control de qualitat de registres existents (14 mencions) és una altra aplicació crítica, demostrant que la IA ja s'utilitza per validar i mantenir la integritat de les bases de dades bibliogràfiques.

Tasques en què s'està emprant la IA	
Assignació de notacions de classificació (CDU, Dewey, LCC, etc.)	23
Creació de registres bibliogràfics complets des de zero (catalogació original)	23
Assignació de capçaleres de matèria, descriptors o paraules clau	20
Generació de resums i notes de contingut	18
Extracció i creació de punts d'accés (autors, entitats)	15
Revisió, correcció i control de qualitat de registres existents	14
Enriquiment de metadades (p. ex., afegint identificadors persistents o enllaços a dades obertes)	11
Conversió/mapeig de metadades entre diferents esquemes	6

7.2.5. Integració de la IA en Models de Metadades

La implementació de la IA en la gestió de metadades mostra una prudència lògica, ja que l'ús d'IA implica treballar directament amb l'estructura de les dades bibliogràfiques. L'estàndard MARC 21 és el model predominant en el qual s'aplica la IA amb 6 mencions, seguit d'RDA amb 9 mencions (si bé el recompte d'RDA és major que MARC 21, MARC 21 sol ser el model base subjacent a nivell tècnic).

Una troballa crucial és que 19 institucions indiquen que, si bé utilitzen la IA, aquesta no s'utilitza per a l'estructuració directa de les metadades, sinó per a altres tasques com l'anàlisi de dades, la interacció amb l'usuari o l'automatització administrativa (com s'infereix de la secció 7.2.).

- **Enfocament Estandarditzat:** L'aplicació en MARC 21 i RDA assenyala que les biblioteques estan integrant la IA dins dels marcs de catalogació existents, la qual cosa és fonamental per assegurar la interoperabilitat i la qualitat de les dades a llarg termini.
- **Enfocament Local:** La menció d'Esquemes de metadades locals/personalitzats (1 menció) i Altre (especifiqueu) (5 mencions) indica una minoria que adapta solucions d'IA a necessitats específiques i no estandarditzades.

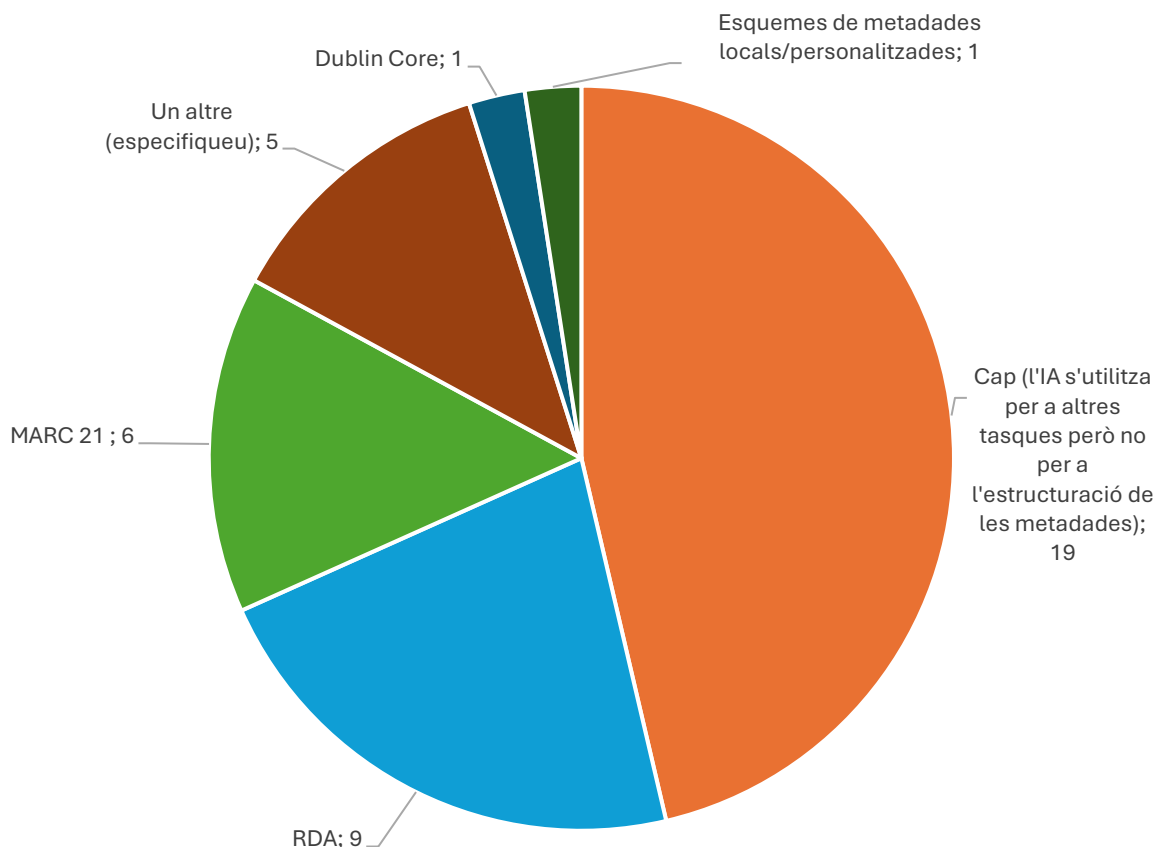


Gráfico 8. Models d'estàndards de metadades en els quals s'usa la IA

7.3. Enfocament específic en catalogació i la seva integració amb la IA

L'ús de la IA en la catalogació és l'aspecte central d'aquest estudi. Per contextualitzar el potencial de la IA, és necessari considerar primer les dades de la catalogació tradicional en les institucions enquestades.

7.3.1. Dades de catalogació anual i proporció manual

La majoria de les institucions enquestades són entitats petites, la qual cosa es reflecteix en els seus volums de catalogació anual:

- Menys de 10.000 registres anuals: 70 institucions.
- La resta dels rangs (10.000 - 100.000 registres) sumen només 14 institucions.

A més del baix volum, la dependència dels mètodes manuals continua sent alta:

- Menys del 25% de catalogació manual: 39 institucions.
- 75%-100% de catalogació manual: 29 institucions.

Això evidencia que gairebé un terç de la mostra manté una alta proporció de catalogació sense protocols d'automatització com Z39.50. La combinació de baix volum de catalogació i alta manualitat constitueix un terreny fèrtil per a l'aplicació de la IA, que podria optimitzar els processos de catalogació original i enriquiment de metadades.

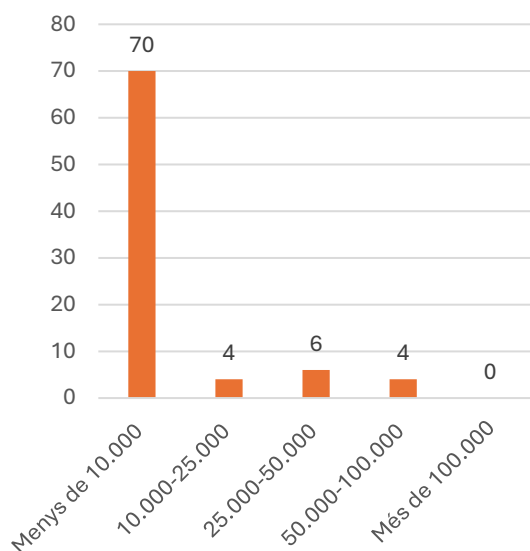


Gráfico 9. Catalogació anual que fa la institució

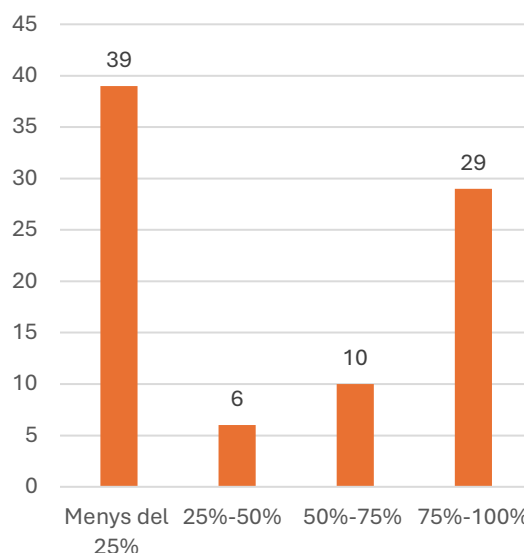


Gráfico 10. % de la catalogació que es fa de forma manual (sense usar protocols com z39.50)

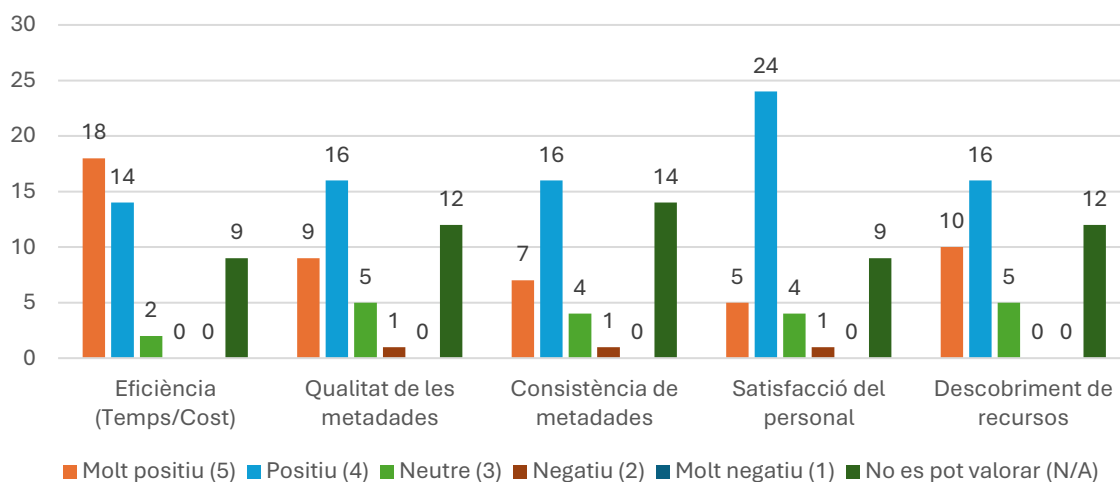
7.3.2. Expectatives d'impacte de la IA en la catalogació

Malgrat el baix ús actiu, les expectatives sobre l'impacte de la IA en la catalogació són aclaparadorament positives

- Eficiència (temps/cost): 32 institucions esperen un impacte positiu o molt positiu (18 Molt Positiu i 14 Positiu). Això indica que la reducció de costos i temps és el principal motor per a l'adopció de la IA en l'àrea de catalogació.
- Qualitat de les metadades: 25 institucions esperen un impacte positiu o molt positiu (9 Molt Positiu i 16 Positiu).
- Consistència de les metadades: 23 institucions esperen un impacte positiu o molt positiu (7 Molt Positiu i 16 Positiu).
- Satisfacció del personal catalogador: 29 institucions esperen un impacte positiu o molt positiu (5 Molt Positiu i 24 Positiu). Això suggereix que la IA es percep com una eina per alleujar la càrrega de treball repetitiu i no com un reemplaçament del catalogador.
- Descobriment de recursos per l'usuari: 26 institucions esperen un impacte positiu o molt positiu (10 Molt Positiu + 16 Positiu).

En totes les categories avaluades, l'impacte positiu o molt positiu supera àmpliament l'impacte neutre o negatiu. El consens és que la IA és un catalitzador per a una catalogació més eficient, consistent i de major valor per a l'usuari final.

	Eficiència (Temps/Cost)	Qualitat de les metadades	Consistència de metadades	Satisfacció del personal	Descobriments de recursos
Molt positiu (5)	18	9	7	5	10
Positiu (4)	14	16	16	24	16
Neutre (3)	2	5	4	4	5
Negatiu (2)	0	1	1	1	0
Molt negatiu (1)	0	0	0	0	0
No es pot valorar (N/A)	9	12	14	9	12



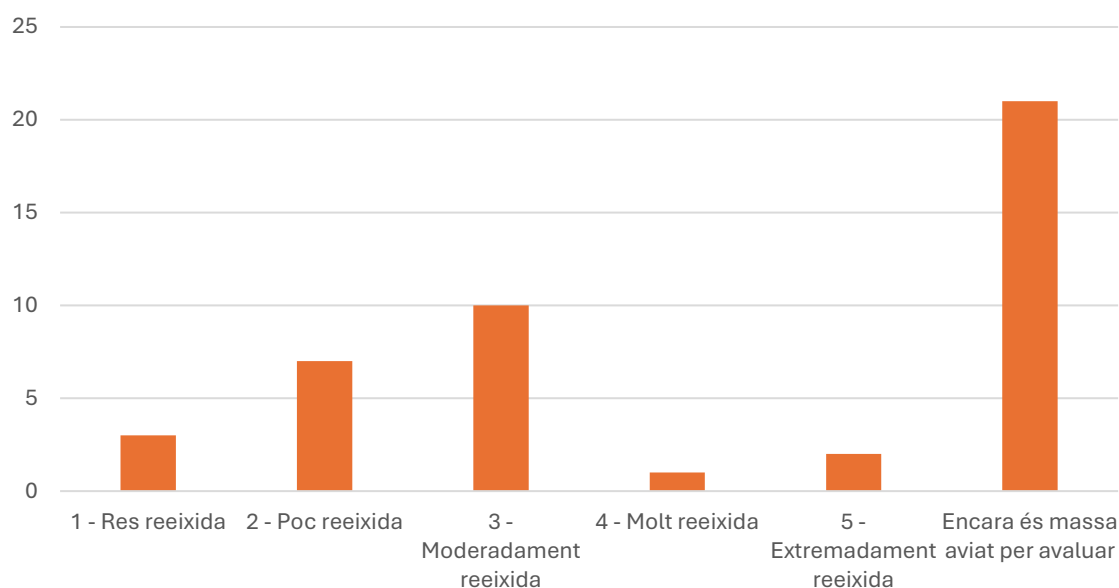
Gràfic 11. Expectatives d'ús de la IA en relació a diferents aspectes de la catalogació

7.3.3. Èxit percebut en la implementació de la IA en catalogació

Per complementar l'anàlisi sobre el nivell d'adopció, es va sol·licitar a les institucions que valoressin el grau d'èxit de la implementació de la IA en la catalogació. Els resultats reforcen la tesi que la tecnologia es troba en una etapa incipient de maduresa en el sector:

- La majoria de les respostes es concentren en l'opció "Encara és massa d'hora per avaluar", amb 21 mencions, la qual cosa subratlla la naturalesa exploratòria o de projecte pilot dels esforços actuals.
- De les 23 institucions que sí que van oferir una valoració sobre l'èxit, la categoria dominant és "Moderadament exitosa" amb 10 mencions.
- Les valoracions entre "Molt exitosa" i "Extremadament exitosa" sumen un total de 3 mencions, mentre que les de "Gens exitosa" i "Poc exitosa" assoleixen les 10 mencions.

Aquest equilibri de percepcions, dominat per una valoració moderada i una gran quantitat d'institucions que encara no poden jutjar, suggereix que els projectes d'IA en catalogació estan oferint resultats preliminars que, si bé prometedors per a alguns, encara no han demostrat un èxit rotund o una transformació disruptiva per a la majoria.



Gràfic 12. Percepcions respecte a la implantació de la IA

En resum, el nivell d'adopció actual de la IA en les institucions bibliotecàries europees es caracteritza per la cautela i la prospecció. La majoria dels enquestats ha superat l'etapa de desconeixement, entrant de ple en la fase d'exploració activa. Malgrat el baix ús productiu actual, existeix un fort consens que la IA tindrà un impacte molt positiu en l'eficiència, la qualitat i la consistència de les metadades de catalogació. L'alta dependència de la catalogació manual i els baixos volums anuals en les institucions enquestades subratllen la necessitat d'eines d'automatització com la IA per transformar un procés que, altrament, consumeix una quantitat significativa de recursos. El baix nombre d'institucions que poden avaluar l'èxit de la IA en catalogació, juntament amb una valoració principalment moderada, confirma que la IA es troba en un estat inicial de desenvolupament i que encara es requereix més temps i projectes pilot per avaluar el seu veritable impacte transformador en el sector.

7.4. Factors organitzatius i desafiaments en la implementació de la Intel·ligència Artificial

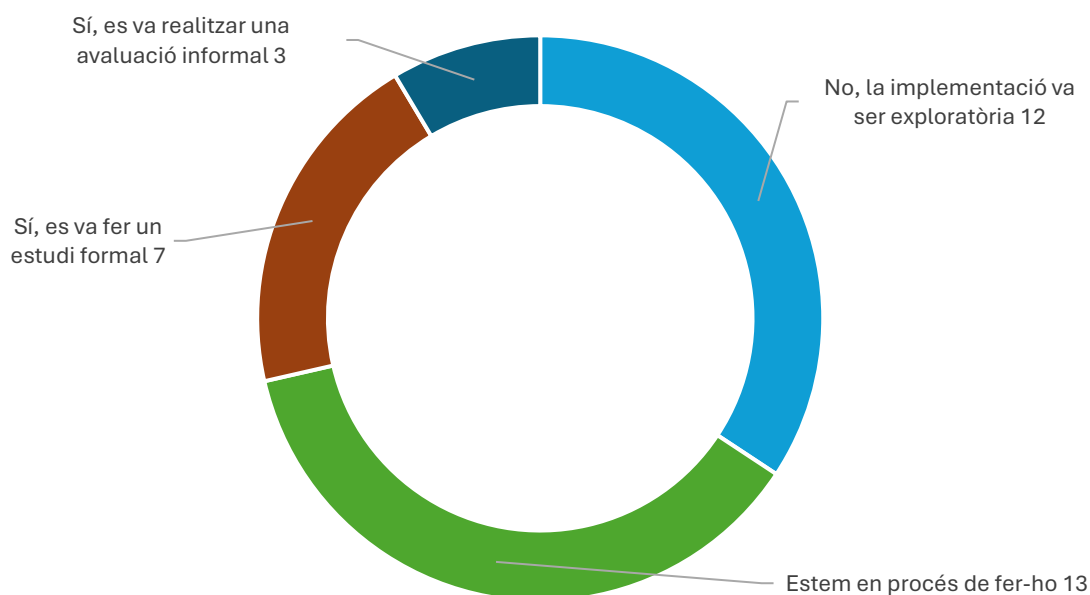
Després d'analitzar el nivell d'adopció actual i els àmbits d'aplicació de la Intel·ligència Artificial (IA) en el sector bibliotecari europeu, és crucial examinar els factors interns que determinen la capacitat de les institucions per gestionar aquest canvi tecnològic. Aquest apartat final se centra en dos elements essencials, sovint ignorats: la solidesa de la planificació estratègica prèvia a la implementació i la composició dels equips humans que lideren els projectes d'IA, evidenciant la naturalesa de les estructures organitzatives que impulsen l'adopció en les biblioteques europees. L'anàlisi d'aquests aspectes és fonamental per comprendre per què la IA es manté majoritàriament en la fase d'exploració en lloc de la producció integrada.

7.4.1. La preparació institucional: desafiaments en la planificació

L'anàlisi de la preparació prèvia a la implementació revela una aproximació cautelosa i, en molts casos, informal per part de les institucions enquestades. Es percep una necessitat de major rigor en la planificació estratègica per a la implementació de la IA. La majoria de les institucions que estan implementant o explorant la IA no han realitzat un estudi formal per avaluar la seva viabilitat o impacte.

- Només 7 institucions afirmen haver realitzat un estudi formal previ a la implementació.
- Un total de 13 institucions es troben en procés de realitzar aquesta anàlisi.
- 12 institucions indiquen que la implementació va ser exploratòria, sense una anàlisi prèvia definida.

Això suggereix que, si bé existeix una etapa d'exploració, els projectes sovint manquen de la fase de planificació estratègica necessària, la qual cosa podria impactar en l'èxit a llarg termini.



Gràfic 13. Nivell d'anàlisi previ a la implementació de la IA

7.4.2. Perfils Clau en la Implantació: La Direcció i el Catalogador

Quant als perfils de personal involucrats, l'estudi posa de manifest que la iniciativa i l'execució es concentren en dos rols principals: la Direcció de la biblioteca i el Personal de catalogació.

- La Direcció de la biblioteca és el perfil més involucrat en la implantació, amb 25 mencions. Això subratlla que l'adopció de la IA es percep com una decisió estratègica que requereix el suport i la visió dels responsables institucionals.

- El Personal de catalogació, amb 23 mencions, és el segon grup més involucrat. Aquesta implicació és fonamental, ja que són els usuaris directes de les noves eines i la seva satisfacció també s'espera que millori amb l'ús de la IA, segons 29 institucions enquestades.
- El Personal del departament de TI / Sistemes (10 mencions) i els Proveïdors de programari externs / Consultors (5 mencions) tenen una representació notablement menor.

Aquesta distribució suggereix que, si bé la decisió i l'aplicació recauen en el personal bibliotecari, la baixa participació dels perfils tècnics interns i externs podria indicar una dependència de solucions amb menor complexitat d'integració o una bretxa en el suport tècnic especialitzat.

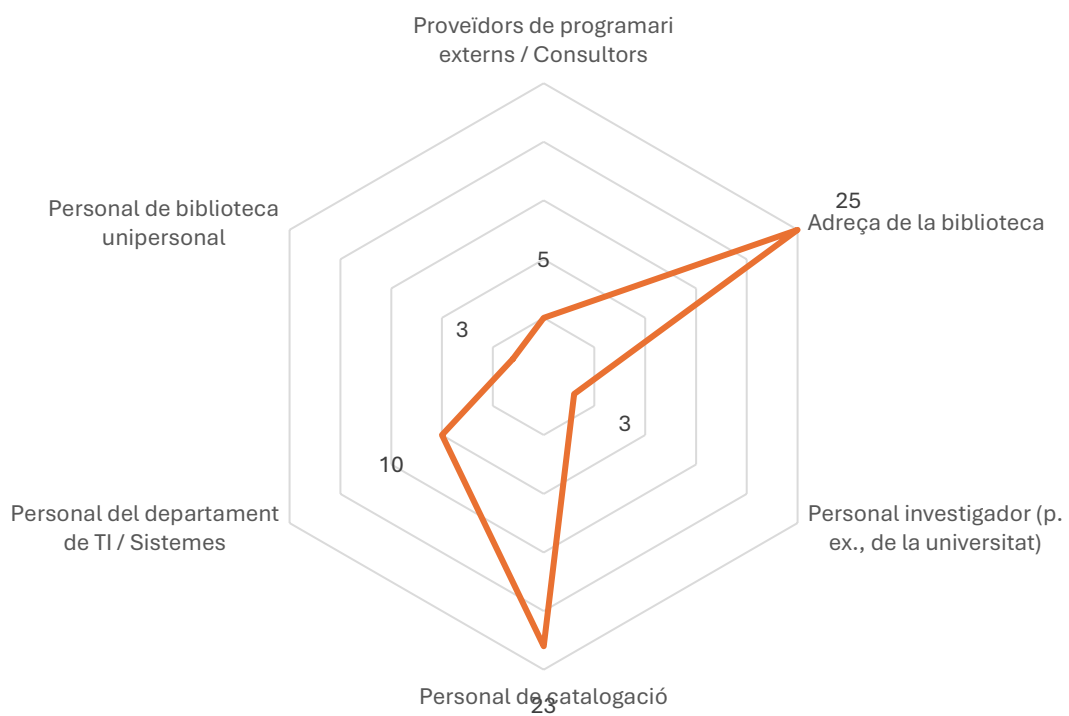


Gráfico 14. Perfils clau en la implantació de la IA

8. Resultats i consideracions finals

L'anàlisi del mercat, la revisió de la literatura científica i les dades de les enquestes institucionals permeten establir un diagnòstic precís sobre l'estat de la intel·ligència artificial aplicada a la catalogació en el període 2020-2025. Com a consideració global cal afirmar que els resultats d'aquest estudi no apunten cap a una substitució tecnològica, sinó cap a una redefinició estructural dels fluxos de treball bibliotecaris.

8.1. El nou rol professional

Existeix un consens en la literatura recent respecte a la inviabilitat de l'automatització total no supervisada. Els resultats confirmen que la IA no reemplaçarà el catalogador, sinó que transformarà el seu rol des de la creació de dades cap a la **supervisió semàntica i l'auditoria**.

Alemu (2025) defineix aquest nou escenari com una col·laboració híbrida, on la tecnologia assumeix la càrrega cognitiva de les tasques repetitives i el reconeixement de patrons a gran escala, mentre que el professional humà aporta tres competències: el judici crític davant l'ambigüitat, la sensibilitat cultural necessària per mitigar biaixos i la validació ètica final. Com conclouen Engel et al. (2025), la supervisió humana (*human-in-the-loop*) és innegociable per garantir la integritat del catàleg i mantenir la confiança de l'usuari en la institució.

8.2. L'equilibri entre sostenibilitat i costos

L'adopció de la IA presenta una dicotomia econòmica. Si bé els models automatitzats prometen una reducció massiva en els temps de processament tècnic, les institucions han de considerar els costos ocults associats a la qualitat de les dades.

La «paradoxa de l'eficiència» detectada en aquest estudi suggereix que la dependència exclusiva de la IA genera registres amb menor densitat informativa, la qual cosa trasllada el cost operatiu a fases posteriors d'enriquiment i correcció manual. Per tant, la sostenibilitat del model no resideix en l'eliminació de personal, sinó en la reassignació de recursos humans cap a tasques d'alt valor afegit, com la gestió d'identitats a la Web semàntica i la curació de col·leccions especials.

8.3. Cap a la convergència: IA neuro-simbòlica

Des d'una perspectiva tecnològica, la investigació identifica que el futur no s'orienta cap a l'elecció exclouent entre IA Generativa o sistemes de regles, sinó en la seva convergència. L'arquitectura **Neuro-simbòlica** emergeix com la solució tècnica que combina la potència lingüística dels Grans Models de Llenguatge (LLMs) amb la fiabilitat estructurada dels Grafs de Coneixement (*Linked Data*).

La implementació d'estratègies com la Generació Augmentada per Recuperació (RAG) permet operar aquesta convergència, mitgant les al·lucinacions en obligar els models a raonar dins dels límits de les ontologies bibliotecàries i els fitxers d'autoritat controlats.

8.4. Recomanacions estratègiques

Basant-nos en l'evidència analitzada, es proposen les següents línies d'actuació per a les institucions bibliotecàries:

- **Alfabetització en IA (AI Literacy).** Es recomana capacitar els catalogadors en enginyeria de *prompts* i en la comprensió de les limitacions algorítmiques, apoderant-los com a auditors crítics del sistema.
- **Adopció d'eines obertes.** Fomentar l'experimentació controlada amb ecosistemes de codi obert (com *Annif* o *OpenRefine*) permet reduir la dependència de l'oligopoli comercial i garanteix una major transparència i control sobre les dades d'entrenament.
- **Col·laboració en xarxa.** La complexitat d'entrenar models propis excedeix la capacitat de la majoria de les biblioteques individuals. En aquest sentit es fa necessari participar en xarxes cooperatives (com els grups de treball de LIBER o REBIUN) per compartir *datasets d'entrenament*, vocabularis controlats i millors pràctiques, assegurant un desenvolupament tecnològic sostenible i alineat amb els valors ètics de la professió.

9. Bibliografía i fonts

Alemu, G. (2025). «Navigating the artificial intelligence frontier on cataloguing and metadata work in libraries: an interview with Getaneh Alemu», *Digital Library Perspectives*, 41(3), pp. 587–592. Disponible a: <https://www.emerald.com/dlp/article/41/3/587/1275650/Navigating-the-artificial-intelligence-frontier-on>

Boté-Vericad, J.-J., Fabeiro, R. i Anglada, R. (2024). *Creando un chatbot con ChatGPT como soporte a la catalogación en bibliotecas, archivos y centros de documentación. Comparación de modelos de lenguaje en versión gratuita y premium*. Barcelona: Universitat de Barcelona. Disponible a: <https://hdl.handle.net/2445/213600>

Breeding, M. (2023). «2023 Library Systems Report: The advance of open systems», *American Libraries*. Disponible a: <https://americanlibrariesmagazine.org/2023/05/01/2023-library-systems-report/>

Breeding, M. (2024). «2024 Library Systems Report», *American Libraries*. Disponible a: <https://americanlibrariesmagazine.org/2024/05/01/2024-library-systems-report/>

Breeding, M. (2025). «2025 Library Systems Report: Companies see platform upgrades, new leadership, and AI enhancements», *American Libraries*. Disponible a: <https://americanlibrariesmagazine.org/2025/05/01/2025-library-systems-report/>

Brzustowicz, R. (2023). «From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging», *Information Technology and Libraries*, 42(3). Disponible a: <https://doi.org/10.5860/ital.v42i3.16295>

Camargo Niño, C. i Zambrano Rivera, S. (2024). «Uso de inteligencia artificial en los sistemas de información documental. Una revisión sistemática de literatura 2013-2023», *Ágora de heterodoxias*, 10(2), pp. 23–46.

CENL. (2024). *AI Survey 2025: analysis*. Conference of European National Libraries. Disponible a: <https://www.cenl.org/wp-content/uploads/2025/05/CENL-AI-survey-analysis-2024.pdf>

Chadwick, B. (2022). «Machine Learning For Classifying Bibliographic Resources: Using topical headings to infer alignments to Australian Curriculum learning areas», *VALA 2022*. Melbourne.

Chisaba-Pereira, C.-A., Herrera-Calero, R., Niño-Neira, S.-A. i Hurtado-Ortiz, B. (2025). «Datalogación: evaluación de herramientas de inteligencia artificial basadas en el Modelo Extenso de Lenguaje (Large Language Model) para la automatización de la descripción de libros», *Infonomy*, 3(4). Disponible a: <https://doi.org/10.3145/infonomy.25.023>

Clarivate. (2025). *Pulse of the Library 2025: Reflecting the Voices of Librarians Worldwide*. Londres: Clarivate Analytics. Disponible: <https://clarivate.com/academia-government/blog/pulse-of-the-library-reflecting-the-voices-of-librarians-worldwide/>

Cotta García, J. G. (2025). *Mejoramiento de la descripción de Recursos Educativos Abiertos, a partir de técnicas basadas en Inteligencia Artificial, Machine Learning y Minería de Datos*. Tesis de Maestría. Bogotá: Universidad Distrital Francisco José de

Caldas. Disponible a: <https://repository.udistrital.edu.co/items/51d7a3ca-d687-44ca-8b20-a6c2ad8dbdca>

Engel, J. Y. et al. (2025) «Artificial Intelligence in Library Cataloging: A Review of Literature», *Journal of Library Metadata*, 25(4), pp. 261–276.

Franganillo, J. (2024). *Inteligencia artificial y bibliotecas: de la euforia a la prudencia*. Presentación en la Asamblea Anual de REBIUN, 13 de noviembre. Disponible a: <https://www.rebiun.org/observatorio-de-inteligencia-artificial>

Gamage, R. C. G. i Wanigasooriya, P. (2024). «Using Generative AI for Bibliographic Description: A Study with ChatGPT 4», *Journal of the University Librarians Association of Sri Lanka*, 27(2), pp. 257-284. Disponible a: <https://jula.sljol.info/articles/10.4038/jula.v27i2.8083>

Golub, K., Suominen, O., Mohammed, A.T., Aagaard, H. i Osterman, O. (2024) «Automated Dewey Decimal Classification of Swedish library metadata using Annif software», *Journal of Documentation*, 80(5), pp. 1057–1079. Disponible a: <https://doi.org/10.1108/JD-01-2022-0026>

Husic, G. (2025). «Leveraging Generative AI for the Creation of Valid MARC21 Bibliographic Records from Unstructured Data: Case Study of Application to Slavic, East European, and Central Asian Language Cataloging», Lawrence, KS: University of Kansas Libraries. Disponible a: <https://hdl.handle.net/1808/36095>

Inkinen, J., Lehtinen, M. i Suominen, O. (2025). «Annif Users Survey: Understanding Usage and Challenges», *Kansalliskirjaston raportteja ja selvityksiä*, 9/2025. Helsinki: National Library of Finland. Disponible a: <https://www.doria.fi/bitstream/handle/10024/190930/Annif%20Users%20Survey.pdf?sequence=1&isAllowed=y>

Jiang, Y., Tang, K., Li, T. i Nakamura, H. (2024). «Exploring the Potential Applications of Generative AI in East Asian Librarianship: Use Cases, Reflections, and Inspirations», *Journal of East Asian Libraries*, 2024, 179(5). Disponible a: <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2866&context=jeal>

Library Technology Guides. (2025). *Library Perceptions 2025: Results of the eighteenth International Survey of Library Automation*. Nashville, TN: Library Technology Guides. Disponible a: <https://librarytechnology.org/pr/31353/library-perceptions-2025-results-of-the-eighteenth-international-survey-of-library-automation>

Lowagie, H. (2023). «One automatic cataloguing flow: tests and first results», *IFLA WLIC 2023*. Rotterdam: IFLA. Róterdam: International Federation of Library Associations and Institutions. Disponible a: <https://repository.ifla.org/items/374e4bf8-bf89-447f-931b-7ac8a7f2ee19>

Martínez-Ávila, D. (2023). «La muerte y el regreso de Bibframe», *Anuario ThinkEPI*, 17. Disponible a: <https://doi.org/10.3145/thinkepi.2023.e17a41>

Miller-Nesbitt, A. (2024). «ChatGPT not Useful as a Tool to Streamline Library Cataloguing Processes», *Evidence Based Library and Information Practice*, 19(2), pp. 145–147. Disponible a: <https://doi.org/10.18438/ebliip30524>

Ng, H. L. i Goh, J. (2025). «SGCAT: Using AI to Facilitate Cataloguing», *IFLA World Library and Information Congress 2025*. Astana: International Federation of Library Associations and Institutions. Disponible a: <https://repository.ifla.org/items/4facbcaf-811e-4181-b3be-e6b85b570aa8>

Ramírez Acosta, A. Á. i Murillo Pabón, J. C. (2024). «Catalogación automática: más cerca que lejos», en *Inteligencia artificial y datos masivos*, pp. 133–148. Laboratorio de Radio e Inteligencia Artificial RTVC. Disponible a: https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/95

REBIUN. (2024). *Inteligencia artificial en bibliotecas universitarias españolas*. Madrid: Red de Bibliotecas Universitarias Españolas (CRUE). Disponible a: <https://www.rebiun.org/noticias/2025/informe-inteligencia-artificial-en-bibliotecas-universitarias-espanolas>

REBIUN. (2025a). *Observatorio de Inteligencia Artificial*. Disponible a: <https://www.rebiun.org/observatorio-de-inteligencia-artificial>

Selbach, C. J., Magnus, A. P. M., Novak, L. D. i Silva, T. da R. (2024). «Transformando as Práticas de Catalogação em Bibliotecas Universitárias: Avaliação do Uso do ChatGPT para o Processamento Técnico na Biblioteca Central da PUCRS», *Biblios*, (87). Disponible a: <https://brapci.inf.br/v/341627>

Storti, C. (2024). «I requisiti per l'addestramento degli strumenti di Intelligenza Artificiale e il deposito legale delle risorse digitali: una riflessione sul contesto normativo italiano e il ruolo delle biblioteche», *Digitalia*, (1), pp. 23–35. Disponible a: <https://doi.org/10.36181/digitalia-00092>

Suominen, O. (2019). «Annif: DIY Automated Subject Indexing Using Multiple Algorithms», *LIBER Quarterly*, 29(1), pp. 1–25. Disponible a: <https://doi.org/10.18352/lq.10285>

Suominen, O. i Koskeniemi, I. (2022). «Annif Analyzer Shootout: Comparing text lemmatization methods for automated subject indexing», *The Code4Lib Journal*, (54). Disponible a: <https://journal.code4lib.org/articles/16719>

Suominen, O., Inkinen, J. i Lehtinen, M. (2022). «Annif and Finto AI: Developing and Implementing Automated Subject Indexing», *JLIS.it*, 13(1). Disponible a: <https://www.jlis.it/index.php/jlis/article/view/437>

Suominen, O., Inkinen, J. i Lehtinen, M. (2025). «Annif at SemEval-2025 Task 5: Traditional XMTTC augmented by LLMs», *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*. Vienna, Austria: Association for Computational Linguistics, pp. 2424–2431. Disponible a: <https://arxiv.org/abs/2504.19675>

Suominen, O., Inkinen, J. i Lehtinen, M. (2025). «Annif at the GermEval-2025 LLMs4Subjects Task: Traditional XMTTC Augmented by Efficient LLMs», *arXiv preprint arXiv:2508.15877*. Disponible a: <https://www.arxiv.org/abs/2508.15877>