

Màster en Arxivística i Gestió Documental

Treball Final de Màster

**L'ÚS DE LA INTEL·LIGÈNCIA ARTIFICIAL ALS
ARXIVS DE LES ADMINISTRACIONS PÚBLIQUES:
UNA REVISIÓ D'ABAST**

Sheila Garcia Reig

Dirigit per Dra. Remei Perpinyà i Dr. Cristóbal Urbano

23 d'agost de 2024

Com citar aquest article: Garcia Reig, Sheila (2024) *L'ús de la intel·ligència artificial als arxius de les administracions públiques: una revisió d'abast*. Treball Final de Màster del Màster d'Arxivística i Gestió de Documents de l'Escola Superior d'Arxivística i Gestió de Documents.. [Http://...](http://...) (consultat el ...)



Es permet la reproducció total o parcial, la distribució, la comunicació pública de l'obra i la creació d'obres derivades, fins i tot amb finalitats comercials, sempre i quan aquestes es distribueixin sota la mateixa llicència que regula l'obra original i es reconegui l'autoria de l'obra original.

RESUM

En aquest treball es presenta una anàlisi contextual sobre el coneixement i desenvolupament de la Intel·ligència Artificial (IA) aplicada a la gestió dels arxius en l'àmbit de les administracions públiques. Aquesta anàlisi s'ha dut a terme mitjançant una revisió sistematitzada, concretament una revisió d'abast o *scoping review*, de la literatura existent. Els resultats obtinguts mitjançant l'anàlisi de 27 estudis seleccionats, juntament amb una revisió del marc teòric, ha permès identificar aspectes clau a tenir en compte sobre el context actual. Finalment, s'han evidenciat sis aspectes rellevants relacionats amb el procés, i vuit relacionats amb el contingut dels estudis. Aquests elements permeten comprendre l'estat actual de l'aplicació i ús de la IA als arxius del sector públic, així com guiar el desenvolupament de solucions eficients per abordar les principals problemàtiques identificades.

Paraules clau: Intel·ligència Artificial, gestió de les dades, arxius, administracions públiques, revisió sistemàtica, revisió d'abast, sector públic.

Título: El uso de la inteligencia artificial en los archivos de las administraciones públicas: una revisión de alcance

RESUMEN

En este trabajo se presenta un análisis contextual sobre el conocimiento y desarrollo de la Inteligencia Artificial (IA) aplicada a la gestión de los datos en el ámbito de las administraciones públicas. Este análisis se ha realizado mediante una revisión sistemática, concretamente una revisión de alcance o *scoping review*, de la literatura existente. Los resultados obtenidos mediante el análisis de 27 estudios seleccionados, junto con una revisión del marco teórico, han permitido identificar aspectos clave a tener en cuenta sobre el contexto actual. Por último, se han evidenciado seis aspectos relevantes relacionados con el proceso, y ocho relacionados con el contenido de los estudios. Estos elementos permiten comprender el estado actual de la aplicación y uso de la IA en los archivos del sector público, así como guiar el desarrollo de soluciones eficientes para abordar las principales problemáticas identificadas.

Palabras clave: Inteligencia Artificial, gestión de los datos, archivos, administraciones públicas, revisión sistemática, revisión de alcance, sector público.

Title: The use of artificial intelligence in government archives: a scoping review

ABSTRACT

This paper presents a contextual analysis of the knowledge and development of Artificial Intelligence (AI) applied to data management in the field of public administrations. This analysis has been carried out by means of a systematic review, specifically a scoping review, of the existing literature. The results obtained from the analysis of 27 selected studies, together with a review of the theoretical framework, allowed us to identify key aspects to be taken into account in the current context. Finally, six relevant aspects related to the process and eight related to the content of the studies were identified. These elements allow us to understand the current state of the application and use of AI in public sector archives, as well as to guide the development of efficient solutions to address the main issues identified.

Keywords: Artificial Intelligence, data management, archives, public administrations, systematic review, scoping review, public sector.

SUMARI

1.	Introducció	5
2.	Estat de la qüestió i marc teòric.....	8
2.1	Dades, dades i més dades. Governança de la informació estructurada	8
2.2	Intel·ligència artificial: redefinint l'abast de la governança de la informació	10
2.3	Gestió documental: un futur d'eines tecnològiques al nostre abast	16
2.4	Context conceptual: La integració de la intel·ligència artificial en la gestió dels arxius de les administracions públiques	22
3.	Metodologia	24
3.1	Tria d'un enfocament de revisió sistematitzada: les <i>scoping reviews</i>	24
3.2	Criteris d'inclusió i exclusió.....	25
3.3	Estratègia de cerca.....	26
3.4	Cribratge i selecció de les fonts.....	29
3.5	Registre de les dades	31
3.6	Síntesi dels resultats i caracterització bibliomètrica	32
4.	Resultats	33
4.1	Anàlisi descriptiva	33
4.2	Anàlisi de l'enfocament.....	39
4.3.	Anàlisi de l'obtenció i la qualitat	43
4.4.	Anàlisi del contingut.....	46
4.4.1	Beneficis i riscos de la IA als arxius de les administracions públiques.....	46
4.4.2	Requisits en l'aplicació de la IA als arxius de les administracions públiques	53
4.4.3	Reptes de futur pel desenvolupament i aplicació de la IA als arxius de les administracions públiques	57
4.4.4	Abast de la recerca sobre l'aplicació de la IA als arxius de les administracions públiques	61
4.5	Limitacions	62
5.	Discussió	64
5.1.	Elements relacionats amb el procés.....	64
5.2	Elements relacionats amb el contingut	67
6.	Conclusions	72
	Declaració de disponibilitat de dades	74
7.	Bibliografia.....	75
	ÍNDIX DE FIGURES.....	86
	ÍNDIX DE TAULES.....	86

1. INTRODUCCIÓ

En els darrers anys la presència de les noves tecnologies en el context de l'anomenada Revolució 4.0, i més concretament la intel·ligència artificial (IA), ha sorgit com una de les eines destinades a transformar el nostre entorn i, per tant, la nostra forma de treballar (Valdiviezo Abad and Bonini 2019). Per aquest motiu, actualment la IA és motiu d'estudi en múltiples disciplines, també en la dels arxius i la gestió documental.

Les aplicacions de IA en la gestió dels arxius són diverses, principalment amb l'objectiu de millorar l'eficiència, la precisió i la seguretat de la informació emmagatzemada, que inclou estratègies que van des del reconeixement òptic de caràcters (OCR) fins a la preservació digital intel·ligent dels arxius (Teel 2024). Tot i això, la major part dels estudis realitzats fins ara s'emmarquen en un àmbit més genèric, com en el cas del projecte *InterPARES I Trust AI (2021-2026)*¹, o en tot cas, centrats en una tipologia d'arxius més de caràcter històric, amb estudis sobre la classificació i descripció de fotografies (Benson 2009, Milleville et al. 2023, Du, Le, and Honig 2024), o aquells que exploren l'ús de l'aprenentatge automàtic per la transcripció automatitzada de documents manuscrits, mitjançant les tècniques de classificació de documents i extracció de metadades (Vidal Ruiz et al. 2020, Couture et al. 2023, Dahl et al. 2023, Tarride et al. 2023).

Un altre àmbit en què l'ús de les TIC, i més concretament l'ús d'algoritmes basats en IA ha augmentat progressivament és el del sector públic, on són cada vegada més presents (Tangi et al. 2023). La principal intenció ha estat la d'aportar múltiples beneficis com ara, la millora de l'eficiència administrativa en el curs de l'activitat diària, així com una prestació de serveis a la ciutadania més efectius i en termes de seguretat pública (Criado 2021, Salvador Serna 2021). Tanmateix, aquesta nova incursió també ha plantejat nous reptes encara per resoldre, emmarcats principalment en termes de privacitat, ètica i equitat.

Davant aquest escenari, el present treball de recerca pretén realitzar una aproximació sobre el coneixement i desenvolupament de la governança i la gestió de les dades en l'àmbit de les administracions públiques, mitjançant les noves tecnologies sorgides recentment, com és el cas de la IA. Si bé és cert que la IA ja és una realitat present en múltiples procediments establerts per les administracions públiques, en l'àmbit arxivístic és una realitat tímidament empresa. Actualment, són aparentment poques les línies de

¹ <https://interparestrustai.org/>

recerca que s'emmarquen estrictament en l'àmbit dels arxius de les administracions públiques, tot i que la gestió documental en l'àmbit administratiu és una de les qüestions més rellevants i que en un futur més podrien beneficiar-se de la IA. Així doncs, i a causa de l'escassa i fragmentada informació que existeix sobre el tema tractat, es pretén identificar quin és el paper actual i futur d'aquesta tipologia d'arxius en la governança de la informació i en relació amb les noves tecnologies com la IA.

Els principals objectius que defineixen aquest treball de recerca són:

- Conèixer l'estat actual sobre el desenvolupament i l'aplicació de la IA als arxius del sector públic, mitjançant l'exploració de les publicacions existents que permeti proporcionar una descripció general completa.
- Evidenciar la importància i amplitud de la relació entre els arxius i les noves tecnologies (concretament la IA) mitjançant una perspectiva analítica i estadística.

D'altra banda, els objectius específics que es pretenen assolir són:

- Explorar les possibilitats de la IA en l'àmbit dels arxius de les administracions públiques i la gestió de les dades.
- Identificar les tendències predominants, així com conèixer les futures línies d'investigació en l'àmbit d'estudi i examinar les possibilitats evolutives d'aquest paradigma.
- Determinar quins són els reptes, requisits i implicacions necessàries que se'ns plantegen com arxivers/es, així com els buits que existeixen al voltant de les investigacions actuals.
- Ser una eina d'utilitat per al coneixement i posterior desenvolupament de solucions eficients per a la seva aplicació pràctica en l'àmbit dels arxius del sector públic, que permetin garantir-ne un ús eficient i responsable, gràcies al coneixement del context actual.

La recerca plantejada pretén elaborar un document de caràcter consultiu per a totes les persones relacionades amb la gestió de la informació en l'àmbit de les administracions públiques, i més concretament per al sector arxivístic, partint d'una aproximació que permeti conèixer les oportunitats i reptes que ens corresponen, així com identificar els possibles buits, però també les possibilitats de millora i/o transformació. La metodologia es fonamenta en una revisió bibliogràfica sistematitzada o *scoping review*, a partir de l'explotació i la posterior anàlisi exhaustiva de les fonts d'informació disponibles, amb la que es pretén realitzar una revisió sistemàtica de la literatura existent sobre el tema.

L'aplicació d'aquesta metodologia s'ha dut a terme seguint el marc de treball més habitual i adient per a les revisions bibliogràfiques sistematitzades, el PRISMA ScR (Tricco et al. 2018). Una tipologia concreta dins del *framework* PRISMA (Tricco et al. 2018, Codina, Lopezosa, and Freixa Font 2021) recomanada davant la voluntat i necessitat de caracteritzar una àrea de coneixement en concret. Aquesta anàlisi es complementarà amb una aproximació bibliomètrica, que permetrà analitzar de manera objectiva, sistemàtica i explícita algunes de les dades, i que conjuntament permetran avaluar des d'una visió integral, quin és l'estat de la qüestió actual, així com extreure'n les conclusions necessàries.

D'altra banda, les fonts d'informació i recerca emprades (bases de dades acadèmiques) han estat escollides de manera justificada i prioritzant aquelles que ofereixen resultats basats en aproximacions sistemàtiques estrictes, com són principalment els articles científics. Tot i això, no hem descartat la recerca en l'anomenada *literatura grisa* així com altres fórmules, davant la necessitat de recopilar la fragmentada informació existent.

El treball s'ha estructurat seguint l'esquema bàsic de l'estil formal i acadèmic, que ha estat pres com a model de referència, modificant, adaptant o ampliant els apartats segons les necessitats i la normativa específica prèviament donada. Així doncs, aquest s'ha estructurat amb els següents títols: un primer apartat d'introducció en el que s'exposa el tema i l'abast de la recerca juntament amb els objectius a assolir. A continuació, un segon apartat on es desenvolupa l'estat de la qüestió a tractar i el marc teòric en què s'emmarca. En el tercer apartat s'explica i justifica la metodologia utilitzada, concretament la revisió sistemàtica de la literatura o *scoping review*. A continuació s'exposen els resultats obtinguts de l'anàlisi dels estudis seleccionats sobre l'ús i aplicació de la IA als arxius del sector públic. Una vegada presentats els resultats, es plantegen una sèrie d'elements a tenir en compte que permeten conèixer quin és l'estat actual sobre el tema tractat, amb l'objectiu de servir com a guia per avançar en el desenvolupament de la IA als arxius, i especialment als arxius públics. Finalment, es descriuen les conclusions extretes d'aquest treball.

2. ESTAT DE LA QÜESTIÓ I MARC TEÒRIC

2.1 Dades, dades i més dades. Governança de la informació estructurada

Actualment les dades són considerades el petroli del s. XXI, una font d'energia explotable i capitalitzable, i com a conseqüència la voluntat de recopilar, emmagatzemar, processar, analitzar i utilitzar aquestes dades és una realitat que va en augment (Casellas Serra 2019, Mordell 2023). Es tracta d'una font inestimable d'informació que recolza la presa de decisions, la innovació i el progrés en una societat globalitzada (Bauman 2001). En conseqüència, la seva importància s'accentuarà a mesura que la quantitat i complexitat d'aquestes dades augmenti en el context digital actual. Per aquest motiu, resulta fonamental l'àmplia comprensió de tot allò que hi té a veure, des de les noves tecnologies (*big data*, intel·ligència artificial, *machine learning*, computació al núvol, etc.) fins a les metodologies i millora de les pràctiques per a la seva gestió (Manyika et al. 2011).

Tot i això, el fet de conservar grans volums de dades no sempre està degudament justificat, principalment si aquesta no reuneix la qualitat informativa suficient. Si definim el concepte d'informació com el "conjunt de dades recollides, processades i emmagatzemades per una organització" (Kampffmeyer 2023, 18), és fonamental la realització d'un procés previ de gestió de les dades, que permeti reduir-ne el volum mitjançant la preservació d'aquella considerada de qualitat, autèntica i confiable (Capellades 2023).

Paral·lelament en els últims anys, el terme governança de la informació ha assolit una gran notorietat en l'àmbit públic, degut a la necessitat de gestionar l'emmagatzematge de grans volums de dades en l'actual era digital. El terme, àmpliament definit, fa referència a les polítiques, els processos i les pràctiques de les organitzacions per gestionar els seus actius d'informació. Això inclou des de la creació, l'emmagatzematge, l'ús i l'eliminació d'informació, així com les polítiques i els procediments que regeixen aquestes activitats que en garanteixin la seguretat, efectivitat i eficiència (Soler Jiménez 2020, Kampffmeyer 2023). Així doncs, el terme denota una implicació en relació amb el domini de la informació com a estratègia global, des del moment en què es crea fins que s'elimina la informació. La norma *ISO 24143:202216 Information and documentation. Information Governance. Concept and principles* és un estàndard internacional que ofereix a les organitzacions una sèrie de criteris pràctics per a l'aplicació efectiva de polítiques i procediments en relació amb la governança de la informació. Entre aquests criteris, les organitzacions han de partir d'una sèrie de principis fonamentals en la

formulació d'aquestes polítiques, procediments i normes: rendició de comptes, integritat, protecció, conformitat, disponibilitat, conservació, eliminació, transparència i seguretat (ARMA 2017, ISO 2022). L'execució d'aquests principis, compromet i assegura a les organitzacions que les seves pràctiques són segures i accessibles, al mateix temps que respecten la privacitat de la informació.

Així doncs, és inevitable afirmar el gran valor que tenen les dades per a les organitzacions, i que progressivament s'està convertint en un actiu essencial. Per aquest motiu, la realització de pràctiques i processos destinats a gestionar la informació estructurada d'una manera eficient i complint amb les necessitats organitzatives i les normatives aplicables, és fonamental per una sòlida governança de la informació (Serra Serra 2024). D'aquesta manera, les organitzacions poden millorar la seva eficiència operativa, prendre decisions més fonamentades, i mantenir la confiança de les diverses parts implicades.

Gestió de les dades a les administracions públiques catalanes

Quan fem referència al sector públic, és important establir un model de govern de dades que garanteixi la gestió eficient de la informació de la qual n'és responsable, amb l'objectiu de garantir la integritat, qualitat, seguretat i disponibilitat de la informació governamental.

En el cas de l'administració catalana, cal destacar el Model de govern de dades, aprovat en l'acord de govern GOV/0158/2023 (DOGC 8967), i en el que s'estableix una gestió que permeti assegurar la correcta transformació digital en què troba immersa actualment, tot assolint una sèrie d'objectius transversals a tota l'administració autonòmica. Per a la seva correcta execució, la normativa estructural ha estat el Decret 76/2020 de 4 d'agost, d'Administració digital (DOGC 8195), que determina els pilars essencials del model, així com els instruments necessaris per implementar-lo.

Amb aquest nou impuls, l'administració catalana pretén generar un canvi estructural que produeixi beneficis tant en l'àmbit intern com en la relació amb la ciutadania. El model es basa en una adaptació de la metodologia internacional DAMA (*Data Management Association*), i es configura al voltant de diversos àmbits estretament relacionats amb la gestió, configuració i preservació de les dades. Un d'aquests àmbits es troba vinculat a la gestió arxivística i preservació de les dades, amb la finalitat d'establir una estructura que n'asseguri la gestió i conservació durant el temps necessari per acomplir la finalitat per la qual s'han generat (DOGC 8967). La responsabilitat principal recau en la Direcció

General d'Administració Digital (DGAD), i tot i que a posteriori es parla de rols més específics com en el cas dels responsables funcionals, en cap cas es fa un esment concret a sectors especialitzats, com en el cas dels/les arxivistes/es. Un aspecte que si trobem àmpliament referenciat en els articles 22-26, del Decret 76/2020 de 4 d'agost d'Administració digital, on es desenvolupa la importància de la gestió arxivística de les dades i dels actius digitals durant tot el cicle de vida dels documents electrònics (DOGC 8195). Aspectes com el disseny, la normalització, la captura, l'ús, la gestió i el tractament, la preservació i l'accés a les dades és, juntament amb l'ús de documents electrònics i altres estructures d'informació digitals, un treball multidisciplinari en què els/les professionals de l'arxivística no només cal que hi participin, sinó que hi tenen un paper molt rellevant (Palomar and Pérez-Ventayol 2023).

Així doncs, la proposta del nou model creat formula un marc comú entorn de la governança de la dada en què es manifesten una sèrie d'ítems fonamentals per al seu futur desenvolupament. Un model dinàmic que caldrà revisar i actualitzar periòdicament en relació amb els canvis operatius i tècnics que pugui experimentar el sector públic en els pròxims anys.

2.2 Intel·ligència artificial: redefinint l'abast de la governança de la informació

El debat entorn els conceptes i les funcions de l'aprenentatge automàtic i la IA ha generat multiplicitat de dubtes. El primer que ens venia al cap quan pensàvem en aquests termes eren ordinadors o màquines amb consciència pròpia, i per tant capaços d'actuar sense la nostra supervisió. Actualment, tots els dubtes i preguntes s'estan resolent, gràcies al desenvolupament d'aquestes noves tendències en àmbits molt propers, com la salut o la relació amb les administracions públiques. La realitat és que tant l'automatització com la IA tenen a veure amb el tractament de grans conjunts de dades, és a dir, amb aquella informació que nosaltres proporcionem al sistema i que abasta des de l'anàlisi i reconeixement de patrons fins a la implementació d'algoritmes d'aprenentatge automàtic en la presa de decisions. Aquesta confluència entre la capacitat de processament de dades i la intel·ligència artificial configura un entorn tecnològic on la sinergia entre l'ésser humà i la màquina és essencial per a la innovació i l'avenç.

L'automatització intel·ligent, també anomenada automatització cognitiva, es fonamenta en la utilització de tecnologies de l'automatització com la IA, la gestió de processos de negoci (BPM) o l'automatització de processos robòtics (RPA) amb la intenció d'optimitzar la presa de decisions de les organitzacions (IBM s.d.). Per aquest motiu,

actualment els processos d'automatització en la gestió de la informació resulten fonamentals quan parlem de la governança de la informació.

Una de les tendències en augment és la de la IA i tot el que inclou, com el processament de llenguatge natural (NLP), la recuperació de la informació intel·ligent o l'aprenentatge automàtic i analític. El terme intel·ligència artificial (IA) fou encunyat l'any 1956 per l'informàtic John McCarthy durant la Conferència de Dartmouth i definit com “la ciència i l'enginyeria de fer màquines intel·ligents” (McCarthy 1956). Tanmateix, durant els anys quaranta els matemàtics N. Wiener i J. V. Neumann ja treballaven en la teoria dels sistemes i la computació, que més tard establiria les bases de la IA (Nofre, Priestley, and Alberts 2014). Cal dir que es tracta d'un conjunt d'eines tecnològiques que existeixen des de fa més de cinquanta anys, però que gràcies a l'avanç de la potència informàtica, així com a la disponibilitat de nous algorismes, han permès realitzar grans progressos en aquest camp durant els últims anys (Criado 2021). Una sèrie de tecnologies innovadores que permetran augmentar l'eficiència, precisió i accessibilitat en la gestió de grans volums de dades, i per tant generar una oportunitat de millora en la governança de la informació.

A continuació, i en relació amb els diferents termes sorgits al voltant de la IA al llarg del treball, es mostra una taula amb els principals conceptes per facilitar la seva comprensió:

Intel·ligència Artificial (IA)	Camp de la informàtica que es dedica a crear sistemes capaços de realitzar tasques que normalment requereixen intel·ligència humana, com l'aprenentatge, el raonament i la percepció.
Automatització de processos intel·ligents (<i>Intelligent Process Automation</i> , IPA)	Conjunt d'eines, basades en tecnologies cognitives derivades de la intel·ligència artificial, que permeten automatitzar processos complexos que requereixen alguna forma de judici o adaptació intel·ligent.
Aprenentatge automàtic (<i>Machine Learning</i>)	Subcamp de la IA que se centra en el desenvolupament d'algorismes i models estadístics que permeten a un sistema realitzar tasques específiques sense ser programats explícitament.
Sistemes de recomanació	Sistemes que utilitzen algorismes per recomanar elements d'interès a l'usuari basant-se en la predicció de les seves preferències i comportament passat.

Reconeixement d'entitats amb nom (<i>Named Entity Recognition</i> , NER)	Procés que permet identificar i classificar automàticament les entitats mencionades en un text en categories predefinides (llocs, persones...).
Processament de llenguatge natural (<i>Natural Language Processing</i> , NLP)	Àrea de la IA centrada a ajudar els ordinadors a comprendre el llenguatge humà, mitjançant l'extracció d'informació específica de text no estructurat.
Aprenentatge profund (<i>Deep Learning</i>)	Subcamp del ML que utilitza xarxes neuronals artificials per modelar patrons complexos en grans volums de dades.
Xarxes neuronals recurrents (<i>Recurrent Neuronal Network</i> , RNN)	Tipus de xarxa neuronal artificial dissenyada per processar dades seqüencials, en què les connexions entre les unitats permeten retroalimentar-se.
Sistemes experts	Sistema informàtic que simula la capacitat del raonament humà davant la presa de decisions.
Raonament basat en casos (<i>Case-Based Reasoning</i> , CBR)	Sistemes que utilitzen casos similars en la base de dades per resoldre problemes nous, tot adaptant les solucions passades a les noves situacions.
Reconeixement òptic de caràcters (<i>Optical Character Recognition</i> , OCR)	Tecnologia que converteix imatges de text imprès o manuscrit en text digital editable.
Xatbot	Programa d'IA que simula una conversa amb usuaris humans mitjançant tècniques de NLP per entendre i respondre a les entrades dels usuaris.

Taula 1. Definició de conceptes entorn la IA. Fonts: (Samoili et al. 2021, Jarvinen 2020, Frankish and Ramsey 2014).

Els canvis tecnològics generats per aquestes eines permetran que els/les professionals de les dades puguin concentrar-se en altres tasques de valor afegit, com l'anàlisi i interpretació de la informació, o l'establiment de les directrius bàsiques per una correcta gestió. La idea principal és que els procediments existents se supervisin, analitzin i millorin a partir de processos d'optimització i recreació automàtica, així com la redefinició de les tasques habituals d'organització i relació amb els/les usuaris/es per mitjà d'aplicacions intel·ligents (Kampffmeyer 2023). Per tant, podem afirmar que l'impacte de les noves tendències tecnològiques apareix com un fet complex amb múltiples aspectes a tractar de manera continuada, tant per al concepte de governança de la informació com per les tasques pròpies dels/les arxivistes/es. Els avantatges que aquestes eines poden proporcionar sota un adequat tractament són fonamentals i, per tant, cal veure-les com una oportunitat per avançar i millorar en les tasques habituals. Tanmateix, és

important que s'implementin de manera ètica i acurada (Dubber, Pasquale, and Das 2020), tenint en compte els possibles biaixos i desigualtats de gènere (Criado-Pérez 2020), racials, de classe... garantint un marc normatiu adequat, i prioritant la privacitat i seguretat de les dades custodiades.

Intel·ligència artificial i administracions públiques

Recentment, el concepte de govern obert ha pres una major rellevància en l'entorn de les administracions públiques. Es tracta d'una cultura de governança que promou els principis de transparència, integritat, rendició de comptes i participació de les parts interessades en suport a la democràcia i el creixement inclusiu (OCDE s.d.). Això inclou una sèrie de polítiques fonamentades en la transparència, les dades obertes i la participació, que tenen com a objectiu millorar la comunicació entre la ciutadania i les administracions. Per assolir aquests principis resulta clau parlar de dades obertes. Un concepte que parteix de la idea que tota aquella informació i documentació gestionada amb recursos públics ha d'estar disponible públicament mitjançant els portals de transparència oficials (Torreblanca Valverde 2021). Per aquest motiu, és important comptar amb un marc jurídic que estableixi les pautes a seguir per una correcta governança de les dades però sense perdre de vista els objectius definits, com en els casos de la Llei catalana 19/2014 de transparència, accés a la informació pública i bon govern (DOGC 6780), o la Llei 19/2013 en l'àmbit estatal (BOE 295).

Per tant, podem afirmar que darrerament les administracions públiques i en conseqüència l'administració electrònica (*e-government*) ha adquirit una major presència en la manera com aquesta es relaciona amb la ciutadania, en un context en què la tecnologia va adquirint més protagonisme. Un increment en el valor de les dades i l'augment progressiu d'aquestes, han estat probablement les principals raons per les quals les administracions públiques han centrat el seu interès en nous models tecnològics que els permetin assolir una governança pública intel·ligent de la informació.

L'interès per la IA en el sector públic ha pres en els darrers anys una rellevància cabdal. Si bé, les administracions han sabut aprofitar el sorgiment de les diverses eines digitals per transformar l'atenció directa a la ciutadania, la seva estructura organitzativa interna, en canvi, s'ha mantingut inalterable en els últims quaranta anys (Ramió 2019). Així doncs, davant la necessitat d'una renovació institucional i organitzativa que contribueixi a una millor representació de les administracions públiques de cara als ciutadans en un context de governança, la IA i altres fórmules tecnològiques han suscitat una gran curiositat en el sector. La capacitat d'aprenentatge automàtic dels sistemes de *machine*

learning, la disponibilitat de grans quantitats de dades, així com la possibilitat de processar dades a un cost cada vegada menor, són probablement algunes de les raons per les quals el sector públic ha posicionat la IA al centre de la seva estratègia (Misuraca and Van Noordt 2020).

Tot i això, i de la mateixa manera que succeeix en la majoria d'àmbits, encara són relativament pocs els estudis que podem trobar relacionats amb l'aplicació pràctica de la IA a les administracions públiques. Fins a l'any 2020 foren comptabilitzades 230 iniciatives per part dels estats membres de la Unió Europea. Entre aquestes, concretament a Espanya, se n'iniciaren un total de 12 (tres en l'àmbit nacional, set en l'àmbit local i dues on participaren conjuntament amb altres països). Els àmbits públics més habituals en els que s'han dut a terme aquests projectes són: els d'ordre públic i seguretat (1), afers econòmics (2), habitatge i serveis comunitaris (1), salut (4), entreteniment (1) i serveis públics generals (3) (Misuraca and Van Noordt 2020). Amb aquestes dades podem afirmar que Espanya està utilitzant la IA en un relatiu nombre de casos i en una àmplia varietat de funcions governamentals, i que es troba entre la mitjana dels països membres de la UE en desenvolupar aplicacions d'IA en l'àmbit del sector públic, tan sols avançat per països com Dinamarca, Estònia, Països Baixos o Portugal, i en els mateixos àmbits d'aplicació que la resta.

Són múltiples els usos de la IA en l'àmbit de les administracions públiques, amb iniciatives que van des de l'adopció de decisions automatitzades fins a la utilització dels sistemes per part dels treballadors públics, amb l'objectiu d'utilitzar-los com a suport en les posteriors decisions administratives. Recentment, l'adopció d'aquestes eines amb l'objectiu d'assistir a les peticions de la ciutadania mitjançant sistemes automatitzats, els anomenats xatbots, són cada vegada més comuns (Androutsopoulou et al. 2019, Berning Prieto 2023). Per tant, i com bé reflecteixen aquestes dades, la majoria d'aplicacions desenvolupades responen a necessitats molt concretes del sector públic com la seguretat (Quijano-Sánchez et al. 2018), la salut (Qian Sun and Medaglia 2019) o la informació, però realment encara són poques les que responen a les necessitats administratives del propi sector, més enllà de la millora en els serveis a la ciutadania. Una d'aquestes tasques pendents és la de la gestió documental administrativa, ja que es tracta d'un dels punts clau en què les administracions públiques cal que centrin els seus esforços per a poder-s'hi beneficiar de manera notable (Berning Prieto 2023). Tot i que les funcions de les quals se'n pot beneficiar són múltiples (classificació automàtica de documents, extracció d'informació, i automatització de fluxos de treball, entre moltes altres), creiem que els treballs realitzats concretament en aquest àmbit són bastant

escassos. Per aquest motiu, aquest treball pretén analitzar mitjançant una revisió de la literatura existent i des d'una perspectiva quantitativa i qualitativa quin és l'escenari actual en aquest àmbit.

Mentrestant, i davant un increment en la presència de la IA en el marc dels governs i les administracions públiques, les polítiques públiques en un context àmpliament tecnològic són un objecte clau que cal establir mentre no sigui massa tard, i que la major part dels governs ja es troba desenvolupant (Valle-Cruz et al. 2020). Cal dir que són molts els articles de referència que han treballat aquest tema (Agarwal 2018, Janssen and Helbig 2018, Birkland 2019, Cerrillo Martínez 2019), amb l'objectiu de conèixer les possibilitats que ens planteja l'adopció d'aquests nous algorismes en l'àmbit públic, però també els riscos i limitacions que comporta així com les possibilitats de resoldre'ls de manera anticipada i eficient.

Concretament, en el marc de la Unió Europea el març de 2023 s'aprovà la primera proposta legislativa sobre IA del món, amb l'objectiu d'establir un estàndard global per a la seva regulació, així com garantir que els sistemes comercialitzats i utilitzats en el marc europeu siguin segurs i respectin els drets i valors fonamentals (Unió Europea s.d.). Però aquesta regulació no es troba aliena a les problemàtiques del sector públic, i en conseqüència, planteja les bases per una òptima transformació que té com a objectiu, millorar els serveis públics oferint una millor atenció a la ciutadania mitjançant diverses iniciatives. En el cas d'Espanya, cal remarcar *l'Estrategia Nacional de Inteligencia Artificial (ENIA)* (Govern d'Espanya 2023), com un dels eixos de *l'Agenda España Digital 2026* i dins del *Plan de Recuperación, Transformación y Resiliencia*. L'objectiu és proporcionar un marc de referència per al desenvolupament d'una IA inclusiva, sostenible i centrada en la millora de les relacions amb la ciutadania. En l'últim informe publicat recentment a través de l'ENIA, es fa referència a les administracions públiques en un dels seus eixos amb l'objectiu de potenciar l'ús de la IA en aquest àmbit. Concretament, fa referència a casos d'ús vinculats a assistents virtuals, lluita contra el frau, prevenció de desastres naturals, ciberseguretat i OCR intel·ligent, amb aplicacions a institucions públiques tan concretes com l'Institut Cervantes o el SEPE. Paral·lelament, també s'esmenta la incorporació robòtica de processos (RPA) per l'automatització de tasques diàries pròpies de l'administració. Finalment, en l'àmbit català també existeix un projecte estratègic propi, *l'Estratègia d'Intel·ligència Artificial de Catalunya - CATALONIA.AI* (Generalitat de Catalunya 2020a). El projecte impulsat pel govern desplega un programa d'actuacions per enfortir la presència de la IA a Catalunya, liderant una generació de coneixement, aplicació social i empresarial, així com la creació

de solucions que fomentin el creixement econòmic i social. Els projectes en els quals participa, es troben centrats a generar eines efectives per al sector empresarial estretament relacionat amb el sector públic. Es tracta d'un dels eixos principals del programa, amb el qual es pretén promoure la incorporació de la IA per a fomentar la innovació de sectors estratègics com la salut i el benestar, el medi ambient, la mobilitat i el turisme entre altres.

Així doncs, resulta positiu que les administracions públiques en tots els seus àmbits regionals apostin per la implementació de sistemes d'IA per la millora dels serveis públics i per a la pròpia gestió interna. Podem afirmar que la major part dels governs són plenament conscients de les seves oportunitats, com la millora de l'eficiència, la productivitat i la presa de decisions, però també de les seves amenaces. I per aquest motiu, l'impuls d'un marc institucional regulador és absolutament necessari en la creació d'un ecosistema de confiança, que tingui presents problemàtiques tan fonamentals com la protecció de dades (Bright et al. 2019), la privacitat (Power 2016) o la desigualtat a causa de la presa de decisions esbiaixada (Pasquale 2015, Criado 2021). Per aquest motiu són fonamentals els estudis realitzats al voltant de l'aplicació de nous algorismes basats en IA en l'entorn de les administracions públiques (Gasser and Almeida 2017, Cerrillo Martínez 2019, Valero Torrijos 2019, Valle-Cruz et al. 2020, Vinuesa et al. 2020, Berning Prieto 2023), i que han permès posar en relleu els beneficis, però també les amenaces, que pot comportar la seva aplicació, i en conseqüència, la necessitat de fer-ho d'una manera responsable i planificada. Per aconseguir-ho, és fonamental la tasca empresa pels governs amb el desenvolupament dels seus marcs reguladors, que denoten una clara aposta per la implementació de la IA d'una manera efectiva i segura.

2.3 Gestió documental: un futur d'eines tecnològiques al nostre abast

La gestió documental, i en conseqüència el paper dels/les arxivers/es, és un component fonamental en la governança de la informació. Al parlar d'aquest terme fem referència a tot el conjunt de polítiques, processos i tecnologies que utilitza una organització per gestionar de manera eficient la informació emmagatzemada, garantint la seva integritat, disponibilitat i confidencialitat, així com el seu correcte compliment normatiu. D'altra banda, la gestió documental s'encarrega principalment d'administrar els documents propis d'aquesta organització al llarg del seu cicle de vida, des de la seva creació fins a la seva disposició final (ISO 2016). Això inclou la seva creació, captura, emmagatzematge, indexació, recuperació, distribució, conservació i eliminació d'una manera eficient i d'acord amb les normatives establertes. Per aquest motiu, sembla força

adient establir una relació directa entre la governança de la informació i la gestió documental, o com resulta més adequat anomenar-ho actualment, gestió de la informació, tot i que sempre des d'una perspectiva arxivística (Casellas Serra 2019, Kampffmeyer 2023, Serra Serra 2024).

Davant aquest nou paradigma, i a causa de l'augment en la creació i compilació de les dades, fa que els arxius es trobin davant la necessitat d'anticipar i sistematitzar com ha de ser la seva recollida (Campos Martínez 2023). Aquests nous conjunts, estructurats en bases de dades, es presenten com la base de tot un entramat arquitectònic amb el qual s'han de poder acomplir les funcions habituals d'una organització. Les seves característiques han de contemplar els amplis volums de tramitació així com la diversitat de formats i productors que hi puguin intercedir, sense oblidar aspectes com la seguretat, integritat i privacitat de les dades emmagatzemades. Per aquest motiu, l'ús de la gestió documental i les noves tecnologies és fonamental com a estratègia en la gestió de documents d'arxiu. Tanmateix, és important tenir en compte que actualment la gestió documental només cobreix, i per tant gestiona, una petita part de tota la informació generada per una organització. Un fet degut principalment, a la complicada tasca de controlar-ne una part important, no considerada documentació d'arxiu (Kampffmeyer 2023).

Les estratègies habituals en la gestió documental es fonamenten en l'aplicació i desenvolupament d'una sèrie de polítiques i procediments, així com en el disseny i implementació d'un sistema adequat a les necessitats de l'organització i que, al mateix temps, compleixi les normatives vigents que li són pròpies (ISO 2016). En conseqüència, els sistemes de gestió documental (SGD) són una eina fonamental en l'administració de les organitzacions, que permeten donar solucions a les necessitats corporatives i en relació amb els/les usuaris/es (Casellas Serra 2019). Últimament, les organitzacions s'han trobat davant la necessitat d'integrar els propis sistemes de gestió documental en sistemes d'informació més amplis. Els motius principals són: un augment en la producció de dades generades; la necessitat de centralització i organització davant la diversitat de tipologies i formats de dades, així com una dispersió d'aquestes en diversos sistemes i plataformes; els requisits d'accés i col·laboració, on resulta essencial que els/les usuaris/es hi puguin accedir fàcilment, però mitjançant un accés segur i col·laboratiu; i finalment, el correcte compliment normatiu i de seguretat per a protegir la informació emmagatzemada. Davant aquestes necessitats, i en l'actual paradigma de les dades, han estat múltiples els organismes que han apostat per la transformació i innovació digital en tots els àmbits. Els nous programes impulsats en l'àmbit europeu (Reglament

2021/694) o l'aprovació del Reglament de Governança de les Dades (DGA) (Reglament 2022/868), han fet palesa la necessitat d'invertir i regular aquests grans conjunts de dades.

Paral·lelament, la immersió de les organitzacions, tant de caràcter públic com privat, en un context cada vegada més digital, ha fet replantejar la tasca dels/les arxivistes/es. Es tracta d'una figura fonamental i integradora, que ha de permetre la realització d'una proposta completa sobre el tractament de les dades, i que faciliti a les organitzacions gestionar de manera eficient la seva informació en un entorn digital en constant evolució. En conseqüència, és absolutament necessari adaptar les funcions pròpies i responsabilitats que li han estat assignades tradicionalment. Garantir l'autenticitat, integritat, fiabilitat, confidencialitat i seguretat de la informació al llarg del temps continua sent la principal premissa de la tasca com arxivistes/es. Un fet que les noves tecnologies són incapaces d'assolir de manera autònoma, sense uns coneixements previs i que són la base del seu mecanisme. Per tant, la participació dels/les arxivistes/es és i serà absolutament necessària per garantir una gestió adequada i efectiva de la informació al llarg del seu cicle de vida.

D'altra banda, cal esmentar el paper de les noves tecnologies com els sistemes distribuïts, l'automatització o la IA en la professió arxivística. Una sèrie d'eines, que utilitzades correctament poden resultar extremadament útils, assegurant la creació i el manteniment de la documentació en un context en què els volums d'informació gestionats per les organitzacions és cada vegada més gran. Segons una investigació, duta a terme l'any 2019 entre els membres d'*ARMA International* i clients de *Cohasset Associates* i *Iron Mountain*, només el 44% dels/les arxivistes/es afirmaven utilitzar "eines automatitzades [...] per localitzar i preservar informació rellevant" (Cohasset and ARMA International 2019). Un resultat que ha variat lleugerament en els últims anys entre els països del nord, principalment degut a l'estandardització de les noves tecnologies en multiplicitat d'àmbits, però que no ha succeït de la mateixa manera en tots els països del sud o més empobrits (Modiba 2022a, Tsabedze 2024). Malgrat tot, resulta rellevant la lleugera reticència entre el col·lectiu, ja sigui derivat pel temor al canvi com per la manca d'inversió de les pròpies organitzacions en un moment en què el volum de documents i informació va en augment (Smallwood 2019).

Intel·ligència artificial i arxius

L'automatització i la IA també estan transformant la gestió dels arxius d'una manera molt significativa i, per aquest motiu, són múltiples els treballs publicats en els darrers anys

(Cameron, Franks, and Hamidzadeh 2023, Cushing and Osti 2023, Colavizza et al. 2022). Algunes de les aplicacions més comunes són la indexació i classificació automàtica, el reconeixement òptic de caràcters (OCR), la cerca avançada, la gestió durant el cicle de vida dels documents com l'eliminació o identificació, la traducció automàtica o aquelles relacionades amb la preservació digital.

Una de les primeres aproximacions en la utilització d'un sistema expert per a la identificació, avaluació, preservació a llarg termini i recuperació de documents digitals fou als anys noranta (Gilliland 2016). Des de llavors s'han dut a terme diferents estudis sobre l'aplicació de la IA als arxius, tot i que sovint basats en l'anàlisi d'eines concretes en un context molt específic (Rolan et al. 2019). Les aplicacions inicials foren relacionades amb la classificació de grans volums de documentació o dirigides a millorar l'accés a la informació mitjançant el reconeixement de text escrit (HTR) (Lee 2018, Goudarouli, Sexton, and Sheridan 2019). Posteriorment, les contínues necessitats del sector i la normalització de les noves tecnologies, motivaren el sorgiment d'altres treballs específics, principalment dirigits a millorar l'eficiència i precisió de tasques com la indexació de documents (Binmakhashen and Mahmoud 2019), la classificació (Shabou et al. 2020) i la cerca i posterior recuperació de la informació (Colavizza, Ehrmann, and Bortoluzzi 2019, Bell 2020, Bocyte and Oomen 2020). Els estudis, que tenen com a principal objectiu millorar l'eficiència operativa, l'accessibilitat i la gestió de la informació al llarg del seu cicle de vida (Aydın 2023, Trehan 2023), es presenten en contextos molt diferents però a la vegada molt concrets, aportant solucions a problemàtiques individuals i, fins i tot, poc extrapolables.

Altres formes en què la IA s'ha integrat en l'àmbit dels arxius ha estat mitjançant tècniques de processament de llenguatge natural (NLP) (Hutchinson 2020), amb mètodes que han permès millorar l'extracció de la informació clau, augmentant la precisió de les indexacions, i per tant millorar la cerca de la informació específica. Un altre dels sistemes utilitzat des d'un inici és el reconeixement òptic de caràcters (OCR), amb el que és possible convertir documents físics en format digital i reconèixer automàticament el text imprès o manuscrit. Aquesta tècnica ha estat molt utilitzada en documentació manuscrita, principalment provinent d'arxius històrics (Chronis and Sundell 2011, Firmani et al. 2017, Mas, Fornes, and Lladós 2016, Fornes et al. 2017), tot i que actualment es continua utilitzant com una part fonamental en el disseny de nous sistemes intel·ligents de gestió documental (Yang et al. 2023).

En tot aquest context, és fonamental esmentar *InterPARES Trust Artificial Intelligence (I Trust AI)*, actualment un dels projectes més importants en el tractament d'aquest àmbit. Concretament, es tracta d'un projecte interdisciplinari multinacional, liderat per la *University of British Columbia*, que té com a objectiu dissenyar, desenvolupar i aprofitar la IA per recolzar la disponibilitat i accés continu a documents públics. Mitjançant la creació d'una xarxa de treball que inclou l'àmbit acadèmic, les institucions arxivístiques i els/les professionals governamentals públics i privats, es pretén reforçar un circuit on es nodreixin totes les parts, enfortint el coneixement i les capacitats de totes elles. Es tracta d'un projecte iniciat l'any 2021 i amb data de fi el 2026, on participen persones i institucions procedents de 30 països de 5 continents diferents, així com 83 universitats, 22 empreses i organitzacions internacionals i 16 arxius nacionals i regionals. El projecte engloba nombrosos estudis específics, liderats per les diferents institucions que hi participen. Cada un d'ells s'adhereix a un dels sis grups de treball establerts segons les diferents funcions arxivístiques: creació i utilització, valoració i adquisició, ordenació i descripció, conservació i preservació, gestió i administració, i, referència i accés. Actualment, el projecte es troba entre la fase 3 (2023-2024) i la fase 4 (2024-2025), que tenen com a objectius principals: establir de quina manera els conceptes i principis d'arxiu poden millorar el desenvolupament d'una IA responsable, i a continuació, validar els resultats obtinguts en la fase anterior mitjançant estudis de cas i casos d'ús, que permetin l'extracció d'una anàlisi detallada dels errors detectats en la seva aplicació.

A causa de la fase en què es troba el projecte, les publicacions en revistes indexades on trobar resultats dels estudis específics en curs, són poques. Tanmateix, són múltiples les conferències, tallers i seminaris en les que es presenten els resultats preliminars dels treballs actuals. Altrament, és important remarcar la transparència del projecte i la contínua actualització de la informació mitjançant la web <https://interparestrustai.org/trust>. Gràcies a la difusió empresa pel projecte podem conèixer i accedir als materials dels diferents estudis. Això ens permet seguir-lo molt de prop, arribant a conèixer els resultats obtinguts fins al moment, així com accedir a materials propis, com ara presentacions i informes que inclouen informació encara no publicada.

Malgrat tots aquests estudis, continuen sent escasses les aplicacions formals de la IA en l'àmbit dels arxius d'una manera més àmplia i heterogènia. Si bé coneixem aplicacions en àmbits i problemàtiques molt concretes, com és el cas dels arxius audiovisuals (Bazán-Gil 2023), els arxius fotogràfics (Proctor and Marciano 2021, Net et al. 2023) o els arxius que contenen documentació històrica (Raventós, Hernández, and

Simon 2023), és important la realització d'estudis que aportin solucions àmplies i aplicables a les diferents tipologies d'arxiu. La utilització de mètodes estàndards i des d'una perspectiva més integral, permetria contribuir a millorar la resolució de problemàtiques comunes i extrapolables a altres tipologies d'arxiu, mitjançant la implementació d'aquestes eines i, per tant, assolint una gestió i administració de la informació més eficient.

En general, la integració de la IA als arxius ofereix moltes oportunitats per a millorar la preservació i l'accés a la informació (Teel 2024), però també planteja importants reptes ètics i pràctics, que cal que siguin tractats de manera rigorosa i reflexiva (Bushey 2023). Alguns d'aquests factors giren al voltant de qüestions tan importants com la privacitat i la confidencialitat de les dades, la transparència o la preservació del patrimoni cultural i social (Strigel Carter et al. 2022, Van der Werf and Van der Werf 2022, Münster et al. 2024). Per aquest motiu, és fonamental l'estudi de casos pràctics, que permetin avançar-nos a les possibles contingències, aportant una solució ràpida i eficient, juntament amb la implicació dels governs i el seu impuls en polítiques públiques de referència.

L'estudi sobre el coneixement i la gestió de riscos o presa de decisions és una qüestió que ha estat tractada àmpliament en l'àmbit dels arxius (Mat Isa and Nordin 2013, Moss and Gollins 2017, Jo and Gebru 2020, Aangenendt 2022, Jaillant 2022a, Jaillant and Caputo 2022, Jaillant and Aske 2024, Casadesús de Mingo 2019, Colavizza et al. 2022, Jaillant and Rees 2023), i que ha estat recolzada operativament per la norma "ISO 31000:2018. Risk Management – Principles and Guidelines on Implementation". De la mateixa manera, el projecte *InterPARES Trust Artificial Intelligence (I Trust AI)*, tracta aquesta qüestió com un dels seus objectius principals, així com una de les tasques del projecte en la seva segona fase (Duranti et al. 2021). L'objecte consisteix a identificar els potencials riscos o limitacions, fet que permet aportar solucions anticipades al problema generat, així com establir una sèrie de principis d'arxiu a seguir per a evitar el sorgiment d'aquestes problemàtiques. En conseqüència, la figura de l'arxiver resulta cabdal en el desenvolupament i l'aplicació responsable d'eines que utilitzin IA. Per aquest motiu, la seva participació juntament amb equip multidisciplinari, és fonamental per a generar solucions òptimes que permetin pal·liar o minimitzar els possibles efectes negatius generats per l'aplicació d'aquests sistemes. La comunicació entre diferents col·lectius professionals, amb els que els/les arxivers/es comparteixen codis d'ètica professional i que en molts casos formen part d'aquest cicle d'arxiu, també és un aspecte a tenir en compte (Jaillant and Rees 2023). L'associació i col·laboració interdisciplinària de professionals com funcionaris/es públics, acadèmics/ques, historiadors/es o

informàtics/ques són essencials per al desenvolupament i l'aplicació d'eines confiabls i ètiques.

2.4 Context conceptual: La integració de la intel·ligència artificial en la gestió dels arxius de les administracions públiques

Davant d'aquest escenari, el marc teòric en què s'emmarca la recerca del treball és el coneixement sobre l'aplicació de les noves tecnologies, concretament la IA, en la gestió i administració de la informació generada en l'àmbit de les administracions públiques, i tot el que implica el seu ús (manteniment, ètica, confiança...), amb la finalitat de garantir la integritat, autenticitat i fiabilitat dels documents en un entorn digital en constant evolució.

Cal entendre la IA com un conjunt d'algoritmes i sistemes informàtics capaços de realitzar tasques pròpies de la intel·ligència humana, com la resolució de problemes, l'aprenentatge, el reconeixement de patrons o la presa de decisions. Per aquest motiu, la recerca s'emmarca tenint en compte tots els seus subcamps, com el *machine learning*, el *deep learning*, el processament de llenguatge natural, la visió per computador o artificial, la robòtica o els sistemes experts, entre altres.

Concretament, la IA ofereix un gran ventall d'oportunitats en l'àmbit de la gestió documental, possibilitant una major eficiència i precisió en la gestió dels documents electrònics, principalment en la seva creació, accés, preservació i ús, així com en la presa de decisions relacionada. Per aquest motiu, és fonamental conèixer quines són les pràctiques realitzades i com aquestes poden afectar en l'autenticitat i integritat dels documents electrònics en un entorn en què la IA té un paper cada vegada més rellevant. En el marc de la gestió dels arxius públics, les eines d'IA també són cada vegada més presents, principalment en la gestió, administració i manteniment d'aquests, amb la finalitat d'obtenir una major eficiència operativa en els processos, així com una millora en la cerca i recuperació de la informació. El reconeixement òptic de caràcters (OCR), la classificació automàtica de documents o l'automatització de processos són algunes de les múltiples eines utilitzades per la millora de processos i extracció de dades en l'àmbit de les administracions públiques. En conseqüència, i gràcies a la possibilitat d'anàlisi de grans quantitats de dades, la IA permet identificar pautes i irregularitats molt útils per a la presa de decisions estratègiques i la millora de les polítiques públiques, així com per garantir el compliment de les regulacions normatives i polítiques de seguretat necessàries.

Per tot això, l'objectiu d'aquesta recerca és principalment la d'unificar, analitzar i sintetitzar la informació disponible sobre aquest àmbit. Es tracta de desenvolupar un document de caràcter consultiu que permeti comprendre les oportunitats i amenaces que se'ns plantegen, així com identificar els possibles buits i oportunitats de millora i transformació.

3. METODOLOGIA

La metodologia emprada es fonamenta principalment en una revisió bibliogràfica sistematitzada o *scoping review*, mitjançant l'explotació i posterior anàlisi exhaustiva de les fonts d'informació disponibles. Els resultats permeten realitzar una revisió d'abast de la literatura existent sobre l'aplicació actual de la IA en l'àmbit dels arxius de les administracions públiques. Addicionalment, es duu a terme una aproximació bibliomètrica que permet analitzar quantitativament alguns dels resultats, amb l'objectiu d'identificar i avaluar tendències i patrons dins del context d'anàlisi.

3.1 Tria d'un enfocament de revisió sistematitzada: les *scoping reviews*

Les revisions sistemàtiques (*systematic reviews*) tenen com a objectiu reunir tota l'evidència empírica que compleix uns criteris establerts prèviament, amb la finalitat de respondre una pregunta concreta (Iberoamericano 2012). Es tracta d'un mètode que té el seu origen en l'àmbit de la salut (Antman et al. 1992, Oxman and Guyatt 1993), però que posteriorment va ser adoptat en altres camps d'estudi com les ciències socials i les humanitats (Kugley et al. 2016); gràcies principalment a la seva validesa científica, contràriament que en el cas de les revisions tradicionals (Booth, Papaioannou, and Sutton 2012, Codina 2024). Existeixen dues categories en el marc de les revisions sistemàtiques: les *systematic reviews* i les *scoping reviews*. Les primeres, amb una base teòrica i de procediment més proper a l'àrea de la salut, són utilitzades sobretot en aquest sector. Les *scoping review* són les més adients quan l'objectiu de cerca està relacionat amb les característiques d'una àrea de coneixement o d'investigació (Arksey and O'Malley 2005, Tricco et al. 2018), i per aquest motiu són les més utilitzades en l'àmbit de les ciències socials.

Aquest treball s'ha realitzat a partir d'aquest model metodològic, on s'ha aplicat una perspectiva estructurada i sistemàtica, que ha permès identificar quins són els estudis més rellevants, la selecció d'aquests i els gràfics de dades que ens permeten conèixer quin és el context sobre el tema tractat. En conseqüència, per garantir la seva validesa, l'aplicació d'aquesta metodologia s'ha dut a terme seguint un dels *frameworks* més habituals i adients per a les revisions d'abast, el PRISMA ScR (Tricco et al. 2018). Una tipologia concreta dins del *framework* PRISMA 2020 (Tricco et al. 2018, Codina, Lopezosa, and Freixa Font 2021), recomanada davant la voluntat i necessitat de caracteritzar una àrea de coneixement en concret.

Per assolir els objectius marcats en aquesta recerca i, per tant, posicionar la investigació feta en relació amb els coneixements existents, s'han formulat una sèrie de preguntes:

- En quin context s'han realitzat investigacions prèvies sobre l'aplicació de la intel·ligència artificial en l'àmbit dels arxius del sector públic (p.e. països, organitzacions...)?
- Quins són els objectius i les aportacions d'aquestes investigacions?
- Quins són els enfocaments o les metodologies (p.e. estudi de cas, revisió sistemàtica...) dutes a terme en aquests estudis sobre l'aplicació de la IA als arxius de les administracions públiques?
- Quins són els àmbits arxivístics d'aplicació on la IA té més presència?
- Quins són els possibles beneficis i riscos d'aquesta aplicació?

3.2 Criteris d'inclusió i exclusió

Les característiques específiques que s'han tingut en compte per a la selecció de l'evidència han estat bastant àmplies, principalment degut a la voluntat de realitzar una recerca exhaustiva sobre l'àrea d'investigació. Per aquest motiu, i en relació amb la tipologia de documents inclosos, s'ha tingut en compte principalment els articles científics provinents de revistes indexades (indiferentment del seu estat de publicació), ja que formen el volum més gran de resultats obtinguts mitjançant la cerca a les principals bases de dades. De totes maneres aquests resultats no eren prou amplis per poder assolir els objectius del treball. Per aquest motiu, s'han inclòs aquells articles procedents de congressos i conferències; així com qualsevol tipus de material audiovisual i comunicacions presentades pels autors/es en aquest mateix context. També s'han tingut en compte els capítols de llibres, articles de revistes no indexades, estudis de cas, guies institucionals o tesis doctorals que tractessin sobre l'àmbit d'interès. En tots els casos s'ha prioritzat aquells estudis que aportessin dades qualitatives enfront de les quantitatives, ja que el principal objectiu de la revisió és identificar i conèixer el context de l'àmbit de recerca, i no exclusivament avaluar quantitativament aquestes dades.

No obstant això, s'ha exclòs de la revisió, per no complir amb els criteris definits prèviament i ser més pròpies de les revisions tradicionals, qualsevol article d'opinió provinent d'un entorn web, obres de pensament o de tipus assagístic, així com les ressenyes de llibres, prioritzant en tot cas la font original. Així mateix, s'han exclòs tots aquells estudis que no es trobessin establerts en bases de dades acadèmiques establertes en entorns web, ja que es tracta d'un àmbit d'estudi i aplicació molt recent, i

que, per tant, s'ha considerat que tota la informació es troba en format digital. D'altra banda, i a causa del gran nombre de resultats obtinguts en el cas concret dels Arxius Nacionals del Regne Unit (TNA), i com que apleguen una gran quantitat de documentació històrica, s'han exclòs totes aquelles referències que no fessin esment estrictament a documentació administrativa. Finalment, cal remarcar que cap estudi ha estat exclòs per raó d'idioma, àmbit geogràfic ni dates.

3.3 Estratègia de cerca

L'estratègia de cerca plantejada de manera transparent i reproducible, té com a objectiu localitzar tant estudis publicats com no publicats. Per aquest motiu s'han utilitzat únicament fonts secundàries, tant bàsiques com especialitzades (p.e. cercadors i bases de dades), ja que permeten accedir a la informació d'una manera ràpida i estructurada. A partir de vocabulari controlat, paraules clau i text lliure s'ha realitzat una primera cerca que ha permès identificar les fonts rellevants en les bases de dades utilitzades, algunes escollides per ser repositoris especialitzats en l'àmbit de recerca: Scopus, *Web of Science* (tenint en compte tota la col·lecció de bases de dades accessibles mitjançant la UAB i FECYT), cercador de la Universitat Autònoma de Barcelona, Catàleg Col·lectiu de les Universitats de Catalunya (CCUC), Dialnet, ÍNDICES-CSIC, *Catálogo Colectivo de la Red de Bibliotecas de los Archivos Estatales* (CCBAE), *Library and Information Science Abstracts* (LISA), Temaria, *E-prints in Library Information Science* (E-LIS). El vocabulari controlat, així com les paraules clau utilitzades en una primera cerca, s'han obtingut de la base de dades terminològica del projecte *InterPARES AI*², i s'han utilitzat tant en català i castellà (segons la base de dades), així com principalment en anglès (Taula 2).

Vocabulari de cerca en el títol / paraules clau

"artificial intelligence", "AI", "machine learning", "deep learning", "records management", "records management data", "managed record", "records keeping", "data governance", "archive", "archival", "archivist", "document management", "public archives", "public records", "public administration", "public sector", "public management", "public governance", "public policy", "management of government", "management of public affairs", "governance", "government", "local government".

Taula 2. Vocabulari utilitzat en la primera cerca. Obtingut de la base de dades terminològica del projecte *InterPARES AI*.

Alguns d'aquests repositoris no han aportat resultats valuosos i per aquest motiu, s'ha realitzat una segona cerca, on s'ha prescindit dels més específics (ÍNDICES-CSIC, LISA, Temaria i E-LIS) i s'ha cercat en altres més amplis com: *OpenAIRE*, *Arxiv*, *ACM Digital Library*, *IEEE Xplore* i *Semantic Scholar*. En cap cas s'han posat límits de data o idioma,

² <https://interparestrustai.org/terminology>

tot i que en alguns casos s'ha restringit la cerca excloent camps d'estudi concrets com la medicina, a causa de la gran quantitat de resultats proporcionats i el seu nul nivell d'interès. Així mateix, s'ha cercat a *Google Scholar* i s'han revisat els resultats de les 10 primeres pàgines, fins que s'ha considerat que ja no aportaven resultats rellevants. També s'han obtingut dos estudis en curs, en format presentació PPT del projecte *InterPARES AI*, publicats de manera oberta a la seva pàgina web. El conjunt de la cerca es va realitzar durant el mes de febrer de 2024, i es pot consultar l'estratègia completa per cada una de les bases de dades utilitzades en la **Taula 3**.

A partir dels estudis útils obtinguts en la primera cerca, s'han extret les paraules clau utilitzades en la segona. Aquesta tasca ha permès definir molt més el vocabulari emprat i, per tant, aconseguir resultats molt més precisos. La utilització de diferents fórmules combinant els diversos operadors booleans a partir de la cerca avançada, així com el conjunt de paraules clau, tant les extretes en un primer moment com les proporcionades pels estudis d'interès, han aportat nous resultats rellevants. Tanmateix, a causa del reduït nombre de fonts útils aconseguides, s'ha decidit completar la cerca mitjançant altres procediments com la revisió dels autors/es, de la bibliografia i de les citacions de cada un dels textos obtinguts, i que es consideren rellevants per a l'objectiu del treball. Aquesta revisió, i la dels resultats proporcionats en les diferents cerques, s'ha realitzat de forma manual, tenint en compte tant el títol com el resum de la font. Tot i això, en alguns casos concrets, ha sigut necessària la lectura del text complet o una part d'aquest, per així poder constatar la necessitat d'incloure'l o no. Tots els estudis o fonts seleccionades s'han introduït al gestor de referències bibliogràfiques *EndNote*³.

Base de dades	Estratègia de cerca
Scopus	TITLE-ABS-KEY ("artificial intelligence" OR "ai") AND TITLE-ABS-KEY ("machine learning" OR "deep learning" OR "records management" OR "records management data" OR "managed record" OR "records keeping" OR "data governance") AND TITLE-ABS-KEY ("archive" OR "public archives" OR "archival" OR "archivist" OR "document management") AND TITLE-ABS-KEY ("public records" OR "public administration" OR "public sector" OR "public management" OR "public governance" OR "public policy" OR "management of government" OR "management of public affairs" OR "governance" OR "government" OR "local government")) (TITLE-ABS-KEY ("artificial intelligence" OR "ai") AND TITLE-ABS-KEY ("archive" OR "public archives" OR "archival" OR "archivist" OR "document management"))

³ <https://endnote.com/>

	AND TITLE-ABS-KEY ("public records" OR "public administration" OR "public sector" OR "public management" OR "public governance" OR "public policy" OR "management of government" OR "management of public affairs" OR "governance" OR "government" OR "local government"))
Web of Science	"artificial intelligence" OR "AI" (Topic) and "machine learning" OR "deep learning" OR "records management" OR "records management data" OR "managed record" OR "records keeping" OR "data governance" (Topic) and "archive" OR "public archives" OR "archival" OR "archivist" OR "document management" (Topic) and "public records" OR "public administration" OR "public sector" OR "public management" OR "public governance" OR "public policy" OR "management of government" OR "management of public affairs" OR "governance" OR "government" OR "local government" (Topic)
OpenAIRE	(public archives AND artificial intelligence) OR (archives AND "artificial intelligence" AND government)
Arxiv	(public archives AND "artificial intelligence") OR (archive/s AND "artificial intelligence" AND government) OR ("artificial intelligence" AND government AND archives) OR ("artificial intelligence" AND public AND "records management") OR ("artificial intelligence" AND government AND records management) OR ("artificial intelligence" AND archives AND government AND records management) OR ("artificial intelligence" AND government AND archives)
ACM Digital Library	[All: "artificial intelligence"] AND [[All: archives] OR [All: "document management"]] AND [All: "records management"] OR [All: "artificial intelligence"] AND [All: archives] AND [All: "records management"] OR [All: "artificial intelligence"] AND [All: archives] AND [All: government] AND [All: "records management"] AND [All: "machine learning"]
IEEE Xplore	("All Metadata": "artificial intelligence") AND ("All Metadata": archives) OR ("All Metadata": "artificial intelligence") AND ("All Metadata": "public archives") OR ("All Metadata": "artificial intelligence" AND "All Metadata": archives AND "All Metadata": government) OR ("All Metadata": "artificial intelligence" AND "All Metadata": archives AND "All Metadata": government AND "All Metadata": "computational")
CCUC	(TITLE-ABS-KEY ("artificial intelligence" OR "ai") AND TITLE-ABS-KEY ("archive" OR "public archives" OR "archival" OR "archivist" OR "document management") AND TITLE-ABS-KEY ("public records" OR "public administration" OR "public sector" OR "public management" OR "public governance" OR "public policy" OR "management of government" OR "management of public affairs" OR "governance" OR "government" OR "local government"))

Catálogo Colectivo de la red de Bibliotecas del CIDA y de los Archivos Estatales (CCBAE)	("inteligencia artificial" AND "archivos administrativos")
Google Scholar	(archives AND "artificial intelligence" AND government)
Semantic Scholar	("artificial intelligence" AND "public archives" AND archives AND "government" AND "records management") OR ("artificial intelligence" AND "public archives" AND archives AND "government") OR ("artificial intelligence" AND "public archives" AND archives AND "government" AND "machine learning")
Cercador UAB	(archives AND "artificial intelligence" AND government)

Taula 3. Estratègies de cerca realitzades en cada repositori/base de dades amb els connectors booleans i paraules clau (ítems: *database / search strategy*).

3.4 Cribratge i selecció de les fonts

Revisió dels títols i resums: En la cerca realitzada a les bases de dades s'han identificat en total 746 referències. Tal com s'identifica a la **Taula 3**, aquesta seria la suma dels resultats obtinguts, tenint en compte la combinació amb el vocabulari i les paraules clau seleccionades que més resultats d'interès ha aportat. El motiu de comptabilitzar-ho d'aquesta manera ha estat perquè moltes de les diferents combinacions utilitzades han aportat resultats ja repetits en altres combinacions. D'altra banda, i provinents de fonts com són els cercadors (Google Scholar, Semantic Scholar i el cercador UAB), s'han identificat 494 referències. Tots els resultats s'han examinat mitjançant una revisió sistemàtica dels títols i els seus resums, per poder valorar el grau d'adequació amb els objectius de la recerca. Per a determinar la seva inclusió s'han utilitzat els següents criteris:

- La utilització de la IA ha de ser part de l'estudi, encara que sigui de forma secundària. Per aquest motiu, també s'han inclòs tots aquells estudis que fan referència al desenvolupament d'una aplicació prèvia que permeti l'execució d'una IA, així com qualsevol altre desenvolupament aplicat en paral·lel o conjuntament.
- La utilització de la IA en l'àmbit dels arxius de les administracions públiques ha de ser l'eix central de l'estudi. En els casos en què aquest se centra en qualsevol

altra tipologia específica d'arxiu (històrics, fotogràfics...) s'ha exclòs de la selecció.

- Les implicacions de la utilització de la IA en la gestió i administració de la informació, així com tot el que té a veure amb el seu ús (manteniment, ètica, confiança...), ha de ser part de l'estudi i vincular-se directament amb els objectius d'aquest. Tots els articles en què l'ús de la IA no està relacionada directament amb els seus objectius, s'han exclòs.

Així doncs, i seguint la mateixa metodologia, s'han exclòs totes les referències duplicades a partir de la inclusió de les fonts considerades pertinents. Per a un correcte cribatge de les referències s'ha seguit l'extensió PRISMA ScR establert per les *scoping reviews* (Tricco et al. 2018).

Revisió dels textos: Durant l'anterior fase de revisió, s'han llegit íntegrament 41 dels textos, per validar si compleixen amb els requisits d'inclusió, ja que el títol i resum en si, no aporten la informació suficient per a determinar-ho. Se n'han descartat 15 per no complir amb els criteris determinats. Finalment, s'han tornat a revisar els textos complets de les 27 referències considerades apropiades per assolir els objectius de la recerca. Aquestes posteriorment han estat validades per un segon revisor amb una àmplia experiència en el sector arxivístic, i considerades adients per complir adequadament amb els criteris d'inclusió.

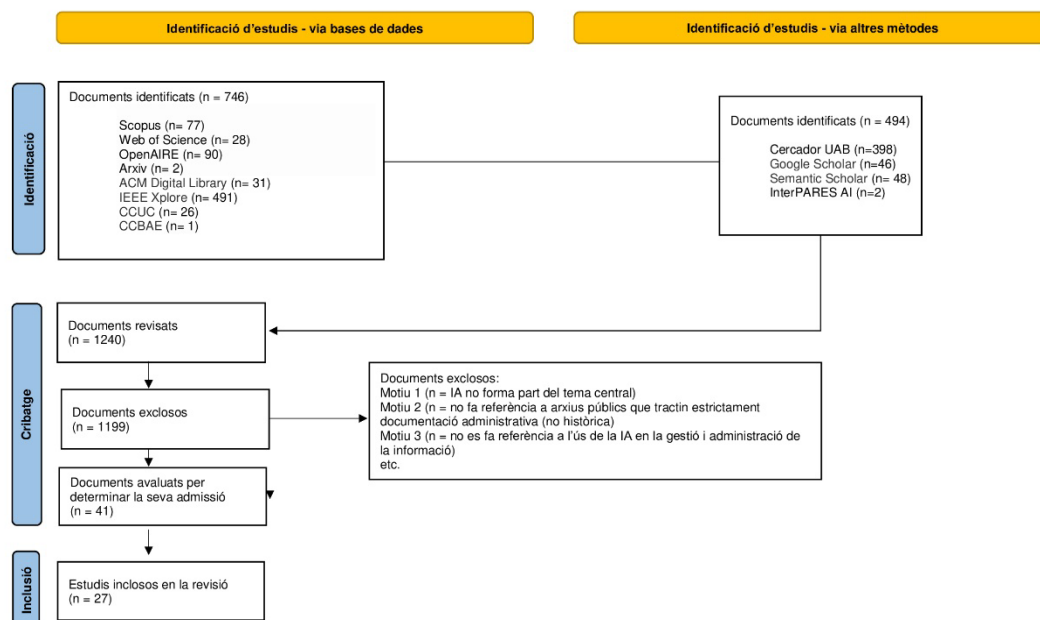


Fig. 1 Selecció, avaliació i inclusió dels estudis (presentats mitjançant el diagrama de flux PRISMA).

3.5 Registre de les dades

El mètode utilitzat per al registre de les dades ha estat a partir d'un esquema prèviament definit en funció dels criteris d'inclusió i exclusió. Aquesta verificació s'ha fet de manera individual, tot i que posteriorment els 41 estudis inclosos han estat revisats i acceptats per un segon revisor. El procés de verificació es pot observar en la **Fig. 1**, on es defineixen les diferents fases dutes a terme (identificació, cribratge i inclusió) per a la incorporació final de les fonts. Els estudis seleccionats s'han introduït al gestor de referències *EndNote*, per una millor administració i explotació de les dades. Posteriorment, també s'han exportat en format CSV per a la seva anàlisi, tenint en compte la informació descriptiva bàsica. A banda, s'han inclòs una sèrie de variables concretes i adaptades als objectius de la recerca, que permeten realitzar la base de l'evidència (**Taula 4**).

Categoria	Metadades	Descripció
Informació descriptiva	# núm. d'article (ID)	número d'estudi assignat al full de treball Excel
	referència completa (APA)	
	any publicació	
	tipus de publicació	(revista, actes d'un congrés, capítol de llibre, tesis...)
	identificador d'objecte digital (DOI)	
	paraules clau	
	base de dades d'obtenció	
	afiliacions autors/es	
Informació sobre l'enfocament	àrea de coneixement autors/es	
	objectiu/s de l'estudi	
	unitat d'anàlisi	unitat específica analitzada (país, organització...)
Informació sobre l'obtenció i la qualitat	metodologia emprada	mètodes utilitzats (quantitatius, qualitatius o mixtes)
	avaluació de la qualitat dels estudis	determinar el correcte disseny i desenvolupament de la investigació, transparència de la informació... (en el cas d'articles provinents de revistes amb

	nombre de citacions	processos <i>peer review</i> , la qualitat de l'estudi pot donar-se per vàlid)
		nre. de citacions en cadascuna de les principals bases de dades utilitzades (<i>Scopus</i> , <i>Web of Science</i> i <i>Google Scholar</i>)
Informació sobre el contingut	resultats (beneficis i riscos)	quins són els beneficis i els riscos de l'aplicació de la IA als arxius públics?
	requisits i desafiaments	requisits i necessitats per mantenir/desenvolupar correctament i de forma sostenible l'aplicació d'IA que s'investiga
	reptes de futur	altres tasques o investigacions necessàries per continuar desenvolupant l'aplicació de la IA en aquest àmbit
	tipus d'IA	(p.e. aprenentatge automàtic, xarxes neuronals, aprenentatge profund...)
	àmbit arxivístic d'aplicació	tipus de tractament de les dades (p.e. gestió i administració, creació i ús, classificació i indexació, preservació, accés...)
	tipus d'organització/institució	organització o institució on s'analitza/aplica

Taula 4. Categories d'anàlisi que han permès la recopilació d'informació per cada estudi seleccionat.

3.6 Síntesi dels resultats i caracterització bibliomètrica

Per presentar la síntesi de resultats s'han utilitzat les diferents categories d'anàlisi determinades anteriorment i agrupades per tipologies (anàlisi descriptiva, d'enfocament, de qualitat i de contingut). Aquestes han permès mesurar individualment cada un dels documents en funció dels seus objectius, resultats, implicacions i àmbits. La informació adquirida i desenvolupada a l'apartat de resultats, ens permetrà produir una síntesi narrativa de caire interpretatiu mitjançant la categorització temàtica dels resultats, evitant únicament la mera descripció, i que permetrà examinar els aspectes principals de l'àmbit d'estudi en relació amb els paradigmes dominants, corrents principals, així com detectar tendències i orientar noves investigacions. A banda, per contextualitzar els resultats s'ha realitzat una breu aproximació bibliomètrica, amb dades dels perfils institucionals i disciplinaris dels autors, així com una petit anàlisi de citacions de la bibliografia emprada en els 27 estudis analitzats.

4. RESULTATS

A continuació, es descriuen els resultats obtinguts a partir de l'anàlisi dels 27 articles seleccionats, que fan referència a l'aplicació de la IA en l'àmbit dels arxius de les administracions públiques (tenint en compte exclusivament aquells que tracten documentació administrativa). Els resultats es mostren categoritzats en les diferents tipologies d'anàlisi determinades a la **Taula 4**, les quals permeten caracteritzar cada un dels documents de forma clara i visible.

4.1 Anàlisi descriptiva

La primera secció de l'anàlisi fa referència a la informació descriptiva. En aquest apartat s'ha inclòs aquella informació bàsica de l'article, com és el cas de l'identificador (ID), la referència completa, l'any i tipus de publicació, el DOI, les paraules clau i la base de dades d'on s'ha obtingut el document (**Taula 5**). Tot i que inicialment la cerca va aportar un registre de 41 documents possiblement vàlids, després d'una revisió exhaustiva del text el nombre d'articles es va reduir a 27, ajustant-se als criteris de selecció establerts prèviament.

ID	Referència + any publicació	Tipus publicació	DOI/ISBN	Paraules clau	Base de dades d'obtenció
# 1	(Hashimoto 2023)	Material audiovisual	-	-	InterPARES AI web
# 2	(Alothman and Sait 2022)	Article de revista	10.1155/2022/4636931	-	Semantic Scholar
# 3	(Azzam et al. 2023)	Article revista	10.3390/app13148463	Blockchain technology Data integrity Documents auditing E-government; OCR	Scopus
# 4	(Alaoui 2023)	Article revista	10.5206/cjils-rcsib.v46i2.16695	Records management Document management Artificial intelligence Teleworking Digital transformation Governance	Web of Science
# 5	(Baron and Payne 2017)	Actes congrés	10.1109/CeDE M.2017.27	Archives E-democracy Electronic records Machine learning Open government Public access Sensitive content	Scopus
# 6	(Chabin 2020)	Article revista	10.1108/RMJ-08-2019-0042	Archival science Artificial intelligence Diplomatics France Great national debate	Scopus
# 7	(Chun 2007)	Actes congrés	1577353234	Case based reasoning	Scopus

				Economic analysis Knowledge based systems Learning systems Project management Automatic assessments Document management systems Immigration agency Immigration application forms Processing systems Electronic commerce	
# 8	(Colavizza et al. 2022)	Actes congrés	10.1145/3479010	Archives Artificial intelligence Machine learning Recordkeeping	Scopus
# 9	(Davet, Hamidzadeh, and Franks 2023)	Article revista	10.1007/s10502-023-09408-8	Accountability Artificial intelligence Automation Interpretability Machine learning Paradata	Scopus
# 10	(Duranti 2022)	Article revista	10.4467/26581264ARC.22.015.16668	Preserving Electronic records Trustworthiness of digital records Project InterPARES Trust AI, Cloud storage Artificial intelligence Blockchain technology	Google Scholar
# 11	(Gelashvili and Pappel 2021)	Actes congrés	10.1109/ICED EG52154.2021.9530905	AI Decision-making Electronic document and records management systems Maturity Model User acceptance Virtual assistants	Scopus
# 12	(Hu et al. 2005)	Actes congrés	10.1002/meet.14504201139	-	Semantic Scholar
# 13	(Ingram and Johnson 2022)	Informe	-	-	-
# 14	(Jaillant 2022b)	Capítol llibre	10.14361/9783839455845	-	Google Scholar
# 15	(Jaillant, Aske, and Caputo 2024)	Informe	-	Digitised archives Born-digital archives Cultural heritage Artificial Intelligence	The National Archives web
# 16	(Jaillant and Rees 2023)	Article revista	10.1093/lc/fqa c073	-	Scopus
# 17	(Lindgren 2022)	Tesis de màster	-	Artificial intelligence AI Transparency EU Public sector	Google Scholar
# 18	(Modiba 2022b)	Article revista	07954778	Artificial intelligence Council for scientific and industrial research Legislative framework Records management	Scopus

# 19	(Modiba, Ngoepe, and Ngulube 2019)	Article revista	10.25159/2663-659X/6159	Records management Archives preservation Artificial intelligence technologies Disruptive technologies	Google Scholar
# 20	(Omigie, Krubu, and Anthony 2023)	Article revista	1743-8278	Archival, Records Management Artificial Intelligence	Google Scholar
# 21	(Reutter and Spiíker 2019)	Capítol llibre	978-383944719-2	-	Scopus
# 22	(Rolan et al. 2019)	Article revista	10.1080/01576895.2018.1502088	-	Scopus
# 23	(Stylianou et al. 2022)	Article revista	10.4018/ijswis.295552	E-government Government portals Linked data Machine learning Natural language processing Open data Semantic web	Web of Science
# 24	(The National Archives 2021)	Informe	-	-	The National Archives web
# 25	(Venkata et al. 2022)	Actes congrés	10.1145/3479008	Artificial intelligence methodologies Document selection Supervised learning methods Document metadata	Scopus
# 26	(Venkata, Young, and Green 2021)	Actes congrés	10.1145/nnnnn nn.nnnnnnn	Text file formats Supervised learning Digital preservation	Google Scholar
# 27	(Wang 2022)	Article revista	10.1111/ssqu.13232	Archive adaptation model Artificial intelligence Information	Scopus

Taula 5. Informació bàsica a nivell descriptiu dels estudis seleccionats.

Tot i que la IA no és un concepte recent des d'un punt de vista acadèmic (McCarthy 1956), ni tampoc és una novetat en el reconeixement de la seva rellevància en l'àmbit dels arxius (Rhoads 1969, Hirtle 1987, Stielow 1991, Benson 2009), la seva aplicació pràctica és un plantejament relativament nou, tant en arxius generals com específicament en arxius públics. Per aquest motiu, es pot observar que els estudis seleccionats pertanyen majoritàriament a una franja de temps molt concreta, compresa principalment en tres anys (2021-2023). Tot i això, el total de la selecció abraça una

franja temporal entre 2007 i 2024, considerant els extrems estadísticament poc rellevants (**Fig. 2**).

Paral·lelament, observem que la major part dels estudis ($n=13$) pertanyen a articles de revistes indexades, així com una part més petita a actes de congressos i conferències ($n=7$). La resta responen a altres tipus de publicacions com són els informes, els capítols de llibre, i de manera molt puntual a materials audiovisuals o

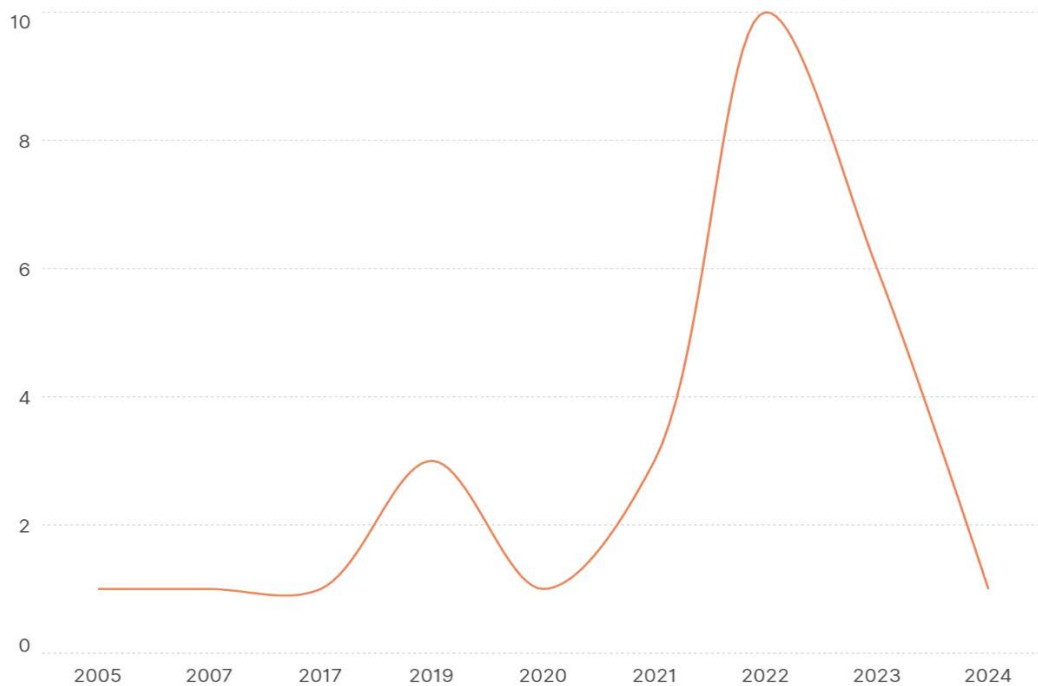


Fig. 2 Gràfic lineal que mostra els anys de publicació dels estudis seleccionats.

treballs acadèmics. En relació amb les bases de dades d'obtenció, majoritàriament s'han extret de *Scopus* ($n=13$), incloent-hi tant articles de revista com actes de congrés. Cal mencionar que una gran part dels estudis també es trobaven a *Web of Science*, però tan sols s'han comptabilitzat en una base de dades per evitar duplicats. En aquesta només s'han comptabilitzat dos estudis que no es trobaven a *Scopus*. Els cercadors *Google Scholar* ($n=6$) i *Semantic Scholar* ($n=2$) han estat molt útils per cercar a partir dels autors/es, bibliografia i citacions dels articles prèviament seleccionats. També s'han extret alguns estudis mitjançant les fonts d'origen, com la pàgina web del projecte *InterPARES AI* ($n=1$) o la dels Arxius Nacionals del Regne Unit (TNA) ($n=2$). Com s'esmenta en l'apartat de metodologia, la cerca en múltiples bases de dades es va realitzar de manera exhaustiva i sistemàtica, per evitar passar per alt estudis significatius. Tot i això, moltes d'elles van aportar resultats duplicats que ja es trobaven

en altres bases de dades científiques més àmplies i amb més possibilitats, com *Scopus* o *Web of Science* (**Fig. 3**).

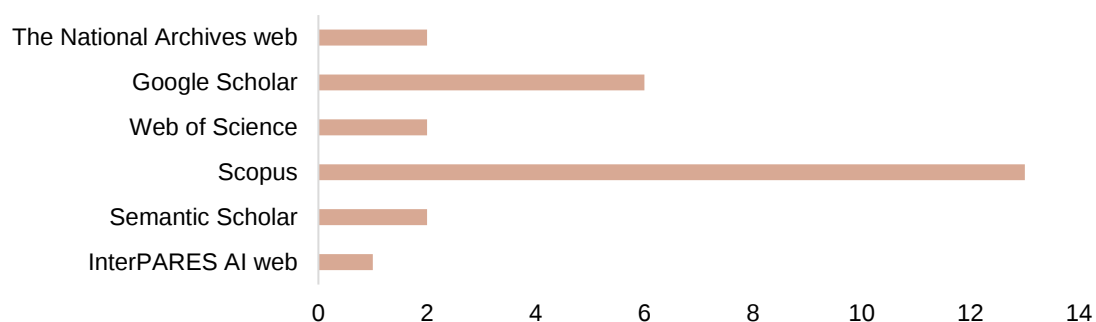


Fig. 3 Bases de dades d'obtenció dels estudis seleccionats.

A continuació, i amb l'objectiu de realitzar una aproximació bibliomètrica, es presenten les dades descriptives referents a les afiliacions dels/les autors/es i el nombre de publicacions en relació amb aquestes. Conjuntament, també es presenta el rang de les àrees de coneixement més representades en la mostra analitzada (**Taula 6**).

Rang	Afiliacions dels/les autors/es	#pub.	Àrea de coneixement	#autors
1	<i>University of British Columbia</i> , Canadà	3	Arxivística	20
2	<i>Loughborough University</i> , Regne Unit	3	Informàtica (Ciències de la Computació)	9
3	<i>The National Archives</i> , Regne Unit	3	Humanitats digitals	4
4	<i>University of South Africa</i> , Sud-àfrica	2	Gestió en patrimoni cultural digital	3
5	<i>Ambrose Alli University</i> , Nigèria	1	Sociologia	1
6	<i>American University</i> , Washington D.C, EUA	1		
7	<i>Aristotle University of Thessaloniki</i> , Grècia	1		
8	<i>Birzeit University</i> , Palestina	1		
9	<i>City University of Hong Kong</i> , Hong Kong	1		
10	<i>Copenhagen University</i> , Dinamarca	1		
11	<i>Department of Finance, Australian Government</i> , Austràlia	1		
12	<i>Digital Archives, NSW State Archives and Records</i> , Austràlia	1		
13	<i>Federal College of Education (Technical) Library</i> , Nigèria	1		
14	<i>International Hellenic University</i> , Grècia	1		

15	<i>King Faisal University, Aràbia Saudita</i>	1
16	<i>Kyoto University, Japó</i>	1
17	<i>Monash University, Austràlia</i>	1
18	<i>National Archives of Australia, Austràlia</i>	1
19	<i>Norwegian University of Science and Technology, Noruega</i>	1
20	<i>Public Record Office Victoria, Austràlia</i>	1
21	<i>San Jose State University, CA, EUA</i>	1
22	<i>Tallinn University of Technology, Estònia</i>	1
23	<i>Université de Laval, Canadà</i>	1
24	<i>Université Paris 8, França</i>	1
25	<i>University of Amsterdam, Països Baixos</i>	1
26	<i>University of Illinois at Urbana-Champaign, IL, EUA</i>	1
27	<i>University of Leeds, Regne Unit</i>	1
28	<i>University of Maryland, MD, EUA</i>	1
29	<i>Uppsala University, Suècia</i>	1
30	<i>Virginia Tech, University Libraries, EUA</i>	1
31	<i>Xihua University, Xina</i>	1

Taula 6. Afiliacions dels/les autors/es més productius/es segons el nombre d'estudis realitzats i àrees de coneixement més representades (#pub: nombre de publicacions, #autors: nombre d'autors).

Els resultats obtinguts mostren que la Universitat de British Columbia (n=3), la Universitat de Loughborough (n=3) i els Arxius Nacionals del Regne Unit (n=3) són les institucions més productives en l'estudi i la recerca de l'àmbit analitzat, seguit de la Universitat de Sud-àfrica (n=2). La resta d'institucions de la llista segueixen un ordre alfabètic, ja que totes apareixien una sola vegada en el total de la mostra.

D'altra banda, i en relació amb l'àrea de coneixement, els resultats mostren una major presència de l'àmbit de l'arxivística (n=20) en el total d'estudis analitzats, seguit d'autors/es que es dediquen a l'àrea de la informàtica o ciències de la computació (n=9). A continuació trobem autors que destaquen per la seva dedicació als àmbits de les anomenades humanitats digitals (n=4) i a la gestió del patrimoni cultural digital (n=3). Finalment, tan sols un dels coautors respon a l'àmbit de la sociologia.

4.2 Anàlisi de l'enfocament

En aquest apartat s'analitzen els diferents aspectes que tenen a veure amb el plantejament dels estudis seleccionats, incloent-hi els objectius (**Taula 7**), la unitat específica analitzada i la metodologia utilitzada.

ID	Referència	Objectiu/s estudi
# 1	(Hashimoto 2023)	1. Conèixer les tecnologies d'IA que estan implementant els governs locals i quins tipus de dades i informació produeixen en la realització d'activitats oficials. 2. Identificar els documents creats per IA i verificar si es gestionen com a "documents administratius". 3. Explorar les implicacions de l'ús d'IA en la gestió de documents públics en l'actual marc legal i administratiu japonès.
# 2	(Alothman and Sait 2022)	Desenvolupar un marc d'IA basat en ontologies per la gestió de documents: 1. Construir un model d'extracció de dades per recuperar text de documents PDF en àrab i anglès. 2. Utilitzar el reconeixement d'entitats nomenades (NER) per classificar documents. 3. Implementar un nou projecte de classificació per recuperar documents rellevants en les consultes d'usuaris.
# 3	(Azzam et al. 2023)	Utilitzar el sistema <i>SECHash</i> (escaneig, extracció, confirmació i <i>hash</i> dels documents), integrat per tecnologies de cadena de blocs i OCR, amb l'objectiu de regular el procés de documents entrants per part d'organismes públics.
# 4	(Alaoui 2023)	Analitzar com la IA pot optimitzar la gestió documental, i els mecanismes de governança que cal desenvolupar per aconseguir-ho.
# 5	(Baron and Payne 2017)	1. Mostrar l'experiència de l'autor en el disseny de polítiques sobre la preservació de documents electrònics del govern dels EUA, com un "cas d'ús" per descriure els reptes que plantegen els "arxius foscos" per la democràcia electrònica. 2. Analitzar les eines i tecnologies emergents, com l'aprenentatge automàtic, per buscar en grans col·leccions de documents emmagatzemats electrònicament i filtrar arxius sensibles. 3. Recomanar una agenda d'investigació i polítiques als organismes governamentals, perquè considerin com a prioritat l'ús de tecnologies de cerca avançada, amb l'objectiu d'aconseguir un equilibri adequat entre accés i privacitat en el tractament de dades confidencials als arxius.
# 6	(Chabin 2020)	Mostrar com l'arxivística pot ajudar a millorar la qualitat dels corpus documentals utilitzats per les eines d'IA i, per tant, a millorar el rendiment d'aquestes.
# 7	(Chun 2007)	Utilitzar la IA per oferir un servei més ràpid i de qualitat a la ciutadania en l'avaluació i processament de les sol·licituds presentades en una oficina d'immigració.
# 8	(Colavizza et al. 2022)	1. Conèixer com es planteja a la literatura acadèmica l'impacte de la IA als arxius institucionals. 2. Conèixer la relació entre la IA i els documents d'arxiu en tot el seu procés (creació, gestió i ús).
# 9	(Davet, Hamidzadeh, and Franks 2023)	1. Considerar el valor i la viabilitat de documentar i preservar els procediments que utilitza la IA, tal com s'aplica a les dades i als documents d'arxiu, promovent així una correcta ètica arxivística. 2. Servir com a punt de partida per a possibles solucions sobre les tensions entre els fluxos de treball en el desenvolupament/implementació d'IA i les particularitats de l'entorn arxivístic.
# 10	(Duranti 2022)	1. Abordar els reptes més importants que afronta l'arxivística en l'era de la transformació digital. 2. Remarcant la necessitat de desenvolupar la teoria arxivística en un món digital, concretament en les àrees d'investigació que requereixen nous desenvolupaments tècnics (aplicació IA).

		3. Remarcar la importància del projecte <i>InterPARES Trust AI</i> , amb l'objectiu de dissenyar, desenvolupar i utilitzar IA per garantir la disponibilitat i l'ús de documents electrònics confiables en entitats públiques.
# 11	(Gelashvili and Pappel 2021)	1. Investigar els reptes que planteja la implementació dels Sistemes de Gestió Electrònica de Documents i Registres (EDRMS) a les entitats públiques georgianes. 2. Determinar el nivell de maduresa actual dels EDRMS i la possibilitat d'implantar assistents virtuals (per IA) en els processos de presa de decisions.
# 12	(Hu et al. 2005)	1. Presentar un projecte real de preservació de la informació governamental en línia a Illinois. 2. Oferir un exemple d'aplicació per a categoritzar text mitjançant l'aprenentatge automàtic semisupervisat en una pràctica real d'arxiu i cerca de publicacions governamentals en línia.
# 13	(Ingram and Johnson 2022)	Idear i articular recomanacions per desenvolupar un projecte pilot que provi i demostrï l'efectivitat de l'aprenentatge automàtic i altres formes d'IA en la creació automàtica de metadades, a partir de les col·leccions digitals de NARA.
# 14	(Jaillant 2022b)	Abordar preguntes sobre l'automatització, accés i IA.
# 15	(Jaillant, Aske, and Caputo 2024)	1. Donar a conèixer el treball actual sobre IA aplicat als arxius, i fomentar futures col·laboracions amb altres institucions. 2. Aportar una perspectiva crítica, des del punt de vista de les Humanitats Digitals i la Informàtica.
# 16	(Jaillant and Rees 2023)	1. Examinar l'aplicació de la IA als arxius digitals com una qüestió que requereix confiança i col·laboració de tot el cercle arxivístic. 2. Entendre els obstacles que s'oposen a l'aplicació d'eines basades en IA per fer més accessibles els fons dels arxius digitals. 3. Identificar les principals preocupacions de les parts interessades en tot el procés arxivístic.
# 17	(Lindgren 2022)	1. Explorar i debatre sobre com garantir la transparència dels sistemes d'IA, i sobre quina informació s'ha de documentar i conservar relativa a ella. 2. Investigar el paper dels arxius i el manteniment dels documents d'arxiu per garantir la transparència de la IA.
# 18	(Modiba 2022b)	1. Investigar com aplicar la legislació actual en l'ús de la IA en la gestió de documents d'arxiu a Sud-àfrica. 2. Proposar un marc per orientar l'aplicació de la legislació en utilitzar IA en la gestió d'arxius.
# 19	(Modiba, Ngoepe, and Ngulube 2019)	Analitzar els riscos en la gestió i conservació dels arxius que poden mitigar-se mitjançant l'aplicació de tecnologies d'IA.
# 20	(Omigie, Krubu, and Anthony 2023)	1. Valorar les perspectives sobre la competència de la IA en els processos de gestió organitzativa. 2. Conèixer el concepte i el domini de la IA a les organitzacions públiques de Nigèria. 3. Determinar els avantatges de l'aplicació de la IA al sistema de gestió d'expedients i arxius (AI-RAMS) dels organismes públics nigerians.
# 21	(Reutter and Spiiker 2019)	1. Analitzar les primeres iniciatives del govern noruec per millorar els serveis públics, en què s'explora una iniciativa per desenvolupar serveis basats en IA. 2. Conèixer els reptes relacionats amb aquest desenvolupament. 3. Examinar les pràctiques reals en l'aplicació de l'aprenentatge automàtic, i els dubtes i reptes metodològics que comporta la utilització de la IA als serveis públics.
# 22	(Rolan et al. 2019)	Oferir una instantània sobre la utilització de la IA en la gestió documental australiana.
# 23	(Stylianou et al. 2022)	1. Presentar Doc2KG (<i>Document-to-KnowledgeGraph</i>), un marc intel·ligent que permet tant la creació com l'actualització en temps real d'un gràfic de coneixement aplicat a sistemes de gestió documental (SGD), al mateix temps que explota estàndards d'ontologia específics. 2. Demostrar els avantatges d'aquest marc a través de la seva aplicació pràctica en un portal real on es publiquen totes les decisions administratives del sector públic, i proporciona accés al SGD governamental on s'emmagatzemen els documents.

# 24	(The National Archives 2021)	Descriure el treball realitzat a TNA, que explora el potencial dels productes comercials basats en eines d'IA per a la gestió del volum, la diversitat i la complexitat dels documents d'arxiu digitals.
# 25	(Venkata et al. 2022)	1. Dissenyar el projecte de TNA, que permet avaluar la idoneïtat i efectivitat de les tecnologies d'aprenentatge automàtic (ML) existents per la selecció i conservació de documents d'arxiu digitals. 2. Reflexionar sobre els reptes a què s'enfronten els gestors documentals, i el paper que pot tenir la tecnologia per abordar-los. 3. Explorar qüestions relacionades amb la confiança i amb la capacitat d'explicar les decisions preses per IA en tasques tan delicades com la selecció.
# 26	(Venkata, Young, and Green 2021)	Investigar mètodes intel·ligents que permetin extraure patrons i informació del contingut dels arxius per identificar formats d'arxiu de text de manera automatitzada.
# 27	(Wang 2022)	Estudiar el desenvolupament d'un sistema d'arxius modern a la Xina i determinar un model prometedor per adaptar-se als reptes globals a l'era de la postveritat.

Taula 7. Resum dels objectius dels estudis seleccionats per la revisió sistemàtica.

En referir-nos a la unitat específica d'anàlisi, examinem quina és la unitat en termes de país o organització. Al parlar sempre d'organitzacions o entitats públiques/governamentals s'ha considerat deixar aquesta metadada dins l'apartat d'anàlisi del contingut, juntament amb l'àrea o àmbit arxivístic d'aplicació. Així doncs, en aquest apartat s'ha fet referència únicament al país d'aplicació de l'estudi, i no pas al país d'origen del/s autor/s ni a la ubicació de la institució que l'ha creat. Tot i això, en molts casos aquestes dades poden coincidir.

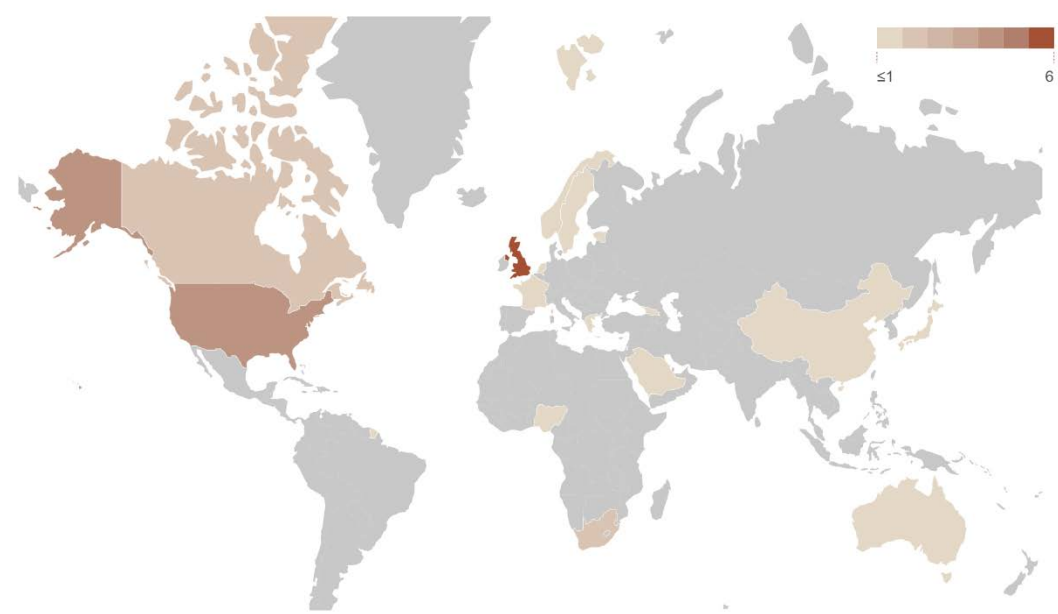


Fig. 4 Mapa coroplètic que mostra els països d'aplicació dels estudis seleccionats, i la quantitat d'estudis duts a terme.

Tenint en compte la **Fig. 4** podem determinar quins són els països més productius en la recerca de l'aplicació d'IA en l'àmbit dels arxius de les administracions públiques. Tal com s'observa, la major part dels territoris compten amb un únic estudi seleccionat (Aràbia Saudita, Austràlia, Estònia, França, Geòrgia, Grècia, Japó, Nigèria, Noruega,

Països Baixos, Suècia i Xina). Cal destacar que hi ha dos territoris més on s'ha dut a terme de la mateixa manera un únic estudi seleccionat, però que el programari no ha detectat, ja que es troben reconeguts de forma limitada, tot i que totalment legítima. Aquests són Palestina i Hong Kong. Seguidament, trobem dos països que compten amb dos estudis cadascun d'ells, que són el Canadà i Sud-àfrica. Els Estats Units d'Amèrica compta amb 4 estudis dels 27 seleccionats, i finalment amb el nombre més alt, el Regne Unit amb un total de 6 estudis.

En referència a la metodologia utilitzada, s'observa que en els estudis de la mostra analitzada s'empraren una gran varietat de mètodes d'investigació, tot i que les revisions bibliogràfiques o de la literatura són, conjuntament amb els estudis de cas, els plantejaments d'investigació dominants ($n=9$). Altres mètodes utilitzats són, la investigació aplicada ($n=5$), les anàlisis de documents oficials i informes (estratègics), els panells d'experts, les revisions sistemàtiques de la literatura, els qüestionaris, les enquestes, les entrevistes i l'avaluació de projectes o iniciatives d'IA existents (**Fig. 5**). És important remarcar que en els estudis seleccionats, la metodologia aplicada no ha de respondre a un únic tipus, sinó que perfectament es poden utilitzar múltiples mètodes d'investigació convergents en un mateix estudi.



Fig. 5 Mètodes d'investigació utilitzats en els estudis seleccionats per la revisió.

En relació amb l'enfocament utilitzat podem dir que els mètodes qualitatius són els més utilitzats en els estudis seleccionats. Gairebé tres quartes parts dels estudis de la mostra són de caire qualitatiu ($n=20$), enfront aquells que utilitzen mètodes mixtos ($n=7$), que combinen enfocaments quantitatius i qualitatius. Particularment, en aquesta mostra no s'ha detectat cap enfocament exclusivament quantitatiu (**Fig. 6**).

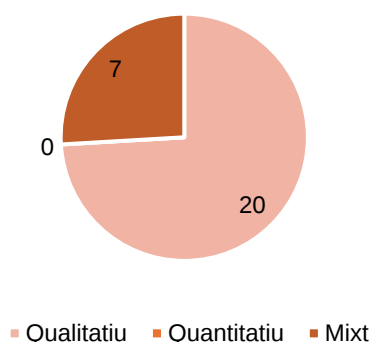


Fig. 6 Enfocament metodològic utilitzat en els estudis seleccionats.

Finalment, cal remarcar que són pràcticament nuls els estudis de la mostra en què les dades de la investigació subjacent es mostren o posen a disposició del lector de manera directa o a través d'un enllaç. Un fet que crida l'atenció, principalment degut a la tendència en augment durant els últims anys, de compartir les dades de la investigació com una pràctica científica oberta, que augmenta

la transparència i la confiança, i que permet l'anàlisi dels resultats (Wilkinson et al. 2016).

4.3. Anàlisi de l'obtenció i la qualitat

En aquest apartat s'analitza i avalua la qualitat dels 27 estudis seleccionats. Sobre el total de la mostra, 15 dels estudis foren obtinguts mitjançant *Scopus* i *Web of Science* i, per tant, s'ha determinat que el disseny i desenvolupament de la investigació és adequat i no planteja cap dubte. El motiu és que són bases de dades que imposen criteris específics i que han de ser aconseguits per qualsevol publicació acadèmica que vulgui ser reconeguda, ja que inclouen elements o indicadors de qualitat indispensables. Aquest filtre assegura als/les usuaris/es un ampli accés a dades de qualitat. D'altra banda, tres dels estudis es van obtenir de la pàgina web de la institució o projecte. Un d'ells (Hashimoto 2023) respon a una presentació en format *PPT*, presentada en una sessió plenària. Tot i això, en tractar-se d'una exposició en què es plantejava la posada en marxa d'un projecte d'equip, i degut a la font de la qual provenia, es va determinar com a vàlida. En el cas de dos informes oficials (The National Archives 2021, Jaillant, Aske, and Caputo 2024), s'han inclòs pel mateix motiu. La resta d'estudis (n=8) es van obtenir mitjançant altres cercadors, i per tant es van revisar detalladament. Tot i això, la majoria dels casos provenien d'editors de revistes científiques *peer reviewed*, que garanteixen uns estàndards de qualitat elevats. A causa de no ser molt estrictes amb els criteris d'exclusió en el moment de la selecció, alguns casos molt puntuals van suscitar certs dubtes, principalment per les fonts d'origen o per la manca d'informació relacionada amb els mètodes d'investigació. No obstant això, cal destacar que alguns d'aquests no responen al tipus de publicació habitual com són els articles de revista (informes, capítols de llibre...), i per aquest motiu els elements o indicadors de qualitat poden ser diferents. Finalment, cal remarcar que els estudis en què la qualitat va

provocar dubtes significatius, ja van ser descartats de la selecció durant la revisió dels textos.

Partint d'aquesta definició de qualitat de la bibliografia científica, s'ha realitzat una explotació de dades en fonts d'informació a partir de les dades bibliomètriques. És per això que es mostra el nombre de citacions dels articles seleccionats en les tres principals bases de dades (*Scopus*, *Web of Science* i *Google Scholar*) (**Fig. 7**). Habitualment, el nombre de citacions que rep una publicació és considerat una mesura fiable del seu valor. No obstant això, en casos on la mostra és relativament petita, com en aquest treball, l'anàlisi de les citacions respon a una consideració més a tenir en compte a l'hora d'avaluar la qualitat dels estudis, tot i que no constitueix un valor absolut per si mateix.

Tot i això, observem que els estudis amb un major nombre de citacions són articles i actes de congressos que es troben correctament indexats a les bases de dades més utilitzades per la comunitat científica. Com s'ha comentat anteriorment, alguns dels estudis inclosos en la mostra pertanyen a altres tipus de publicacions com, capítols de llibre, informes... un fet que dificulta la seva cerca i posterior recuperació d'una manera ràpida i senzilla. En aquest cas no podem atribuir una relació directa entre el nombre de citacions i l'any de publicació dels estudis. En aquest cas observem que pràcticament tots són estudis relativament recents, però, en canvi, mostren un nombre de citacions molt desigual i que responen a altres factors.

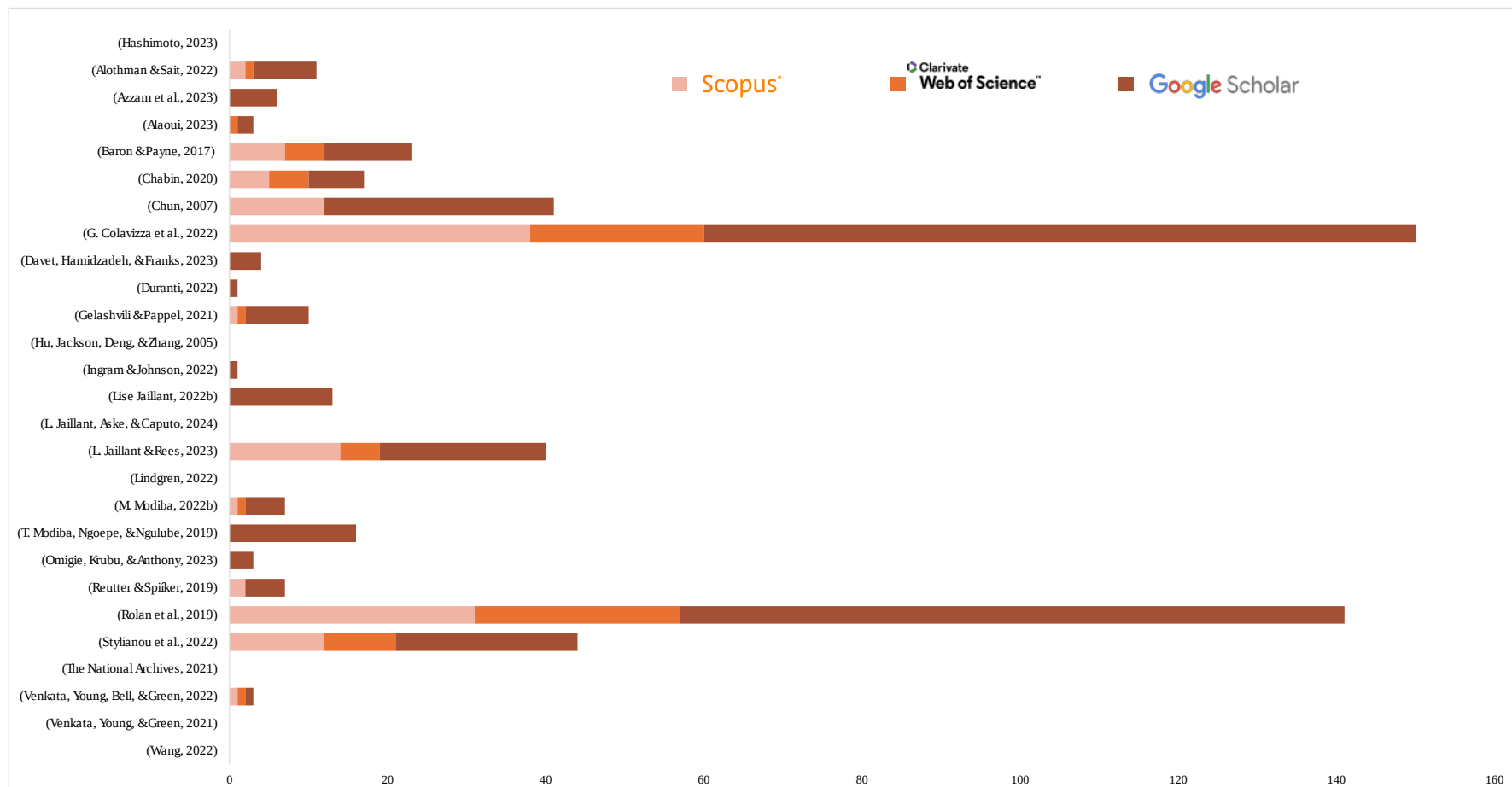


Fig. 7 Nombre de citations en les principals bases de dades utilitzades.

4.4. Anàlisi del contingut

En aquesta secció, es detalla tota aquella informació dels estudis seleccionats que té relació amb el seu contingut, com els beneficis i riscos en l'aplicació de la IA (4.4.1), els requisits necessaris (4.4.2) i els reptes de futur (4.4.3). A continuació es presenta una descripció de l'abast dels articles seleccionats (4.4.4), en què s'inclou el tipus d'IA utilitzada en l'estudi, l'àmbit arxivístic d'aplicació i el tipus d'organització / institució en què s'analitza o aplica.

4.4.1 Beneficis i riscos de la IA als arxius de les administracions públiques

En aquest apartat s'analitzen els beneficis i riscos que comporta l'aplicació i ús de la IA en l'entorn dels arxius de les administracions públiques, a partir de la identificació de resultats dels estudis seleccionats. Cal remarcar que només s'han tingut en compte els arguments que pertanyien als resultats dels/les autors/es i no a altres fonts citades. Per a una millor comprensió, els beneficis s'han identificat mitjançant 10 categories creades a partir de la lectura del conjunt dels articles (**Taula 8**).

Categoria	Beneficis
1) Organització i classificació	Permet noves formes d'organització més fluides (Colavizza et al. 2022) Accelerar el procés de classificació / desclassificació gràcies a la presa de decisions automatitzada (Ingram and Johnson 2022) Permet avaluar, classificar i eliminar la documentació de forma més eficient (Rolan et al. 2019) Redueix el treball manual en l'encapçalament de matèria (Hu et al. 2005) Classificar documents segons el seu nivell de sensibilitat (Jaillant, Aske, and Caputo 2024) Permet destacar temes subjacents entre un gran conjunt de dades gràcies a l'aprenentatge automàtic (Jaillant, Aske, and Caputo 2024) Identificar termes i metadades rellevants (Modiba, Ngoepe, and Ngulube 2019) Crear metadades automàtiques (Ingram and Johnson 2022) Identificar relacions entre entitats i dades (Modiba, Ngoepe, and Ngulube 2019) Identificar entitats relacionades i evita duplicats (Wang 2022) Millorar la classificació minimitzant les pèrdues de documents (Modiba, Ngoepe, and Ngulube 2019) Capacitat de personalitzar documents d'arxiu mitjançant la recopilació i construcció d'un llenguatge algorítmic (Omigie, Krubu, and Anthony 2023)
2) Cerca i recuperació de la informació	Ofereix precisió, recuperació i exactitud (Alothman and Sait 2022) Ajudar en les tasques de selecció de documents de col·leccions semiestructurades i no estructurades (The National Archives 2021) Millorar l'eficiència de la cerca i recuperació de la informació electrònica (Hu et al. 2005) Recuperar documentació que li manca una descripció acurada a nivell d'objecte (Ingram and Johnson 2022) Recuperar documents més ràpidament (Modiba, Ngoepe, and Ngulube 2019) Millorar els processos de recuperació i refinar les cerques (Stylianou et al. 2022) Recuperar arxius a qui els falten metadades descriptives mitjançant el reconeixement facial (Ingram and Johnson 2022) Establir sistemes de recomanació que ajudin a la cerca (Ingram and Johnson 2022) Millora de les metadades mitjançant OCR (Ingram and Johnson 2022) Pot ajudar a recuperar material d'arxiu (Jaillant, Aske, and Caputo 2024) Permet extraure grans quantitats de dades que es troben dins els arxius governamentals (Reutter and Spieker 2019)

	Exacte i precís en la identificació de formats d'arxiu (Venkata, Young, and Green 2021) Permet reduir la càrrega manual en els processos de valoració i selecció de grans volums de documents d'arxiu (Venkata et al. 2022) Permet eliminar deficiències en la recuperació de documents (Omigie, Krubu, and Anthony 2023)
3) Gestió i administració	Millorar la gestió d'expedients gràcies a l'aprenentatge automàtic (The National Archives 2021) Millora en la gestió de la documentació (Gelashvili and Pappel 2021) Facilitar la gestió de la creixent quantitat de dades digitals (Lindgren 2022) Augmentar la precisió, la confiabilitat, l'autoritat i la innovació en els processos de gestió (Omigie, Krubu, and Anthony 2023) Ajuda a millorar la gestió de documents (Modiba 2022b) Alleugerir problemes identificats relacionats amb la gestió i preservació de documents (Modiba, Ngoepe, and Ngulube 2019) Millorar la capacitat de gestió (Rolan et al. 2019) Millorar l'avaluació i selecció en la transferència de documents digitals (Lindgren 2022) Entorn eficaç en la gestió de documents (Allothman and Sait 2022) Permet la construcció i disposició d'un sistema intern de gestió de documents (Omigie et al., 2023) Modernitzar els sistemes de gestió documental (SGD) (Stylianou et al., 2022) Ajuda els arxius a superar les limitacions dels arxius tradicionals (Colavizza et al. 2022) Millora les eines ja existents i permet crear-ne de noves tenint en compte les necessitats dels arxius (Duranti 2022) Transforma el paper de l'arxiu en la gestió de les dades: de receptor passiu a proveïdor actiu (Reutter and Spilker 2019) Major democratització i inclusió en la recopilació i anàlisi de dades (Chabin 2020) Promou noves formes d'arxiu més democràtiques i inclusives (Colavizza et al. 2022)
4) Seguretat i protecció de les dades	L'aprenentatge automàtic permet desenvolupar eines per identificar contingut inusual, ocult o potencialment privat (Baron and Payne 2017) Identificar automàticament arxius confidencials (Jaillant 2022b) Permet detectar la presència d'informació sensible i protegida (Rolan et al. 2019) Capacitat per identificar, catalogar, codificar, indexar documents de manera segura (Omigie, Krubu, and Anthony 2023) Col·labora amb la possibilitat de divulgar informació considerada no confidencial (Jaillant, Aske, and Caputo 2024) Permet controlar l'accés il·legal i no autoritzat a la informació (Modiba 2022b) Capacitat de detectar i prevenir perills relacionats amb la protecció de dades (Omigie, Krubu, and Anthony 2023)
5) Accés	Facilita l'accés a grans volums de documents (Omigie, Krubu, and Anthony 2023) Permet un accés als documents més efectiu (Modiba, Ngoepe, and Ngulube 2019) Millora la qualitat i accés a la informació (Gelashvili and Pappel 2021) Arxius més accessibles (Jaillant 2022b) Facilita l'accés i ús dels arxius digitals (Jaillant and Rees 2023) Gestió dels permisos d'accés basant-se en rols, comportaments i necessitats (Baron and Payne 2017) Resol problemes d'accés a arxius digitals (Jaillant, Aske, and Caputo 2024)
6) Automatització de processos i presa de decisions	Optimitzar la gestió documental (Alaoui 2023) Automatitzar parts dels fluxos de treball d'arxiu (Colavizza et al. 2022) Contribueix positivament en el procés de presa de decisions (Jaillant and Rees 2023) Major transparència i responsabilitat (Gelashvili and Pappel 2021) Beneficiar als arxius en tasques recurrents (Lindgren 2022) Procés de presa de decisions més ràpid i amb menys pèrdua de dades (Gelashvili and Pappel 2021) Redueix els errors humans i entrada de dades incorrectes (Azzam et al. 2023) Flexibilitat que permet millorar la seva capacitat d'aprenentatge fins a obtenir el model adequat (Venkata, Young, and Green 2021) Capacitat de millora constant gràcies a l'aprenentatge continu (Wang 2022)

	Millora la capacitat de processar grans quantitats de dades (Baron and Payne 2017) Processar grans quantitats de dades més ràpidament (Jaillant 2022b) Possibilitat de filtrar i processar grans volums de dades (Venkata et al. 2022)
7) Conservació i manteniment	Garantir el correcte emmagatzematge de documents (Modiba, Ngoepe, and Ngulube 2019) Tecnologia estable i fàcil d'utilitzar per a l'emmagatzematge i preservació dels documents (Omigie, Krubu, and Anthony 2023) Resol problemes de preservació (Jaillant, Aske, and Caputo 2024) Digitalització més ràpida i eficaç (Modiba, Ngoepe, and Ngulube 2019) Accelera els processos de conversió de documents analògics a digital (Modiba, Ngoepe, and Ngulube 2019) Permet reagrupar documentació dispersa, fer-la localitzable i utilitzable (Jaillant 2022b)
8) Compliment normatiu	Assegurar el compliment de les lleis i regulacions actualitzades (Chun 2007) Permet realitzar auditories de forma més automàtica i eficient (Modiba, Ngoepe, and Ngulube 2019) Promou una major garantia en els controls de qualitat, amb possibilitat d'auditar i gestionar sistemàticament els processos de valoració de la documentació (Omigie, Krubu, and Anthony 2023) Permet millorar els processos de rendició de comptes d'una organització (Stylianou et al. 2022) Garantir l'autenticitat dels documents (Jaillant 2022b) Permet comprovar la veracitat de la informació (Wang 2022) Els models generats permeten adaptar els arxius als reptes globals en l'era de la postveritat (Wang 2022)
9) Econòmics	Major rendibilitat operativa (Omigie, Krubu, and Anthony 2023) Capacitat d'aprenentatge continu, sense necessitat d'intervenció humana (Modiba, Ngoepe, and Ngulube 2019) Bona relació cost – eficàcia (Gelashvili and Pappel 2021) Reducció de costos (Azzam et al. 2023) Generar més oportunitats laborals entre els joves, reduint els problemes de desocupació (Omigie, Krubu, and Anthony 2023) Facilita el creixement econòmic del sector públic (Chun 2007)
10) Relació amb el servei i els usuaris	Millora en el temps de resposta del servei, sense afectar-ne la qualitat (Baron and Payne 2017) Millora de la productivitat i qualitat del servei (Chun 2007) Més agilització en els processos i fluxos de treball (Chun 2007) Minimitzar l'error humà en les tasques diàries (Omigie, Krubu, and Anthony 2023) Permet treballar més ràpid i dur a terme tasques monòtones amb més exactitud i precisió (Omigie, Krubu, and Anthony 2023) Oferir millors serveis i més personalitzats (Reutter and Spilker 2019) Satisfer les necessitats d'una base d'usuaris més variada (Colavizza et al. 2022) Millorar l'experiència dels usuaris (Ingram and Johnson 2022) Millora les relacions entre l'administració i la ciutadania (Chun 2007) Interacció directa amb l'usuari (Omigie, Krubu, and Anthony 2023) Accés complet a la informació governamental en línia (Hu et al. 2005) Reduir la càrrega de treball dels empleats públics (Stylianou et al. 2022)

Taula 8. Beneficis de l'ús de la IA en els arxius públics a partir dels resultats dels estudis seleccionats.

En primer lloc, els beneficis relacionats amb l'organització i classificació són, una major fluïdesa, rapidesa i eficiència en l'avaluació, classificació i eliminació de la documentació (Rolan et al. 2019, Colavizza et al. 2022, Ingram and Johnson 2022) gràcies a la presa de decisions automatitzada. Permet reduir el treball manual (Hu et al. 2005) i millorar la classificació minimitzant les pèrdues (Modiba, Ngoepe, and Ngulube 2019). Gràcies al seu nivell de sensibilitat a l'hora de classificar documents (Jaillant, Aske, and Caputo 2024), permet identificar temes i metadades rellevants de manera automàtica (Modiba,

Ngoepe, and Ngulube 2019, Ingram and Johnson 2022, Jaillant, Aske, and Caputo 2024), així com identificar entitats relacionades evitant duplicats (Wang 2022).

En segon lloc, els beneficis en relació amb la cerca i recuperació de la informació són principalment una millora en la precisió, eficiència i rapidesa de les tasques (Hu et al. 2005, Modiba, Ngoepe, and Ngulube 2019, The National Archives 2021, Alothman and Sait 2022, Ingram and Johnson 2022, Stylianou et al. 2022) gràcies a eines com el reconeixement facial, el reconeixement òptic de caràcters (OCR) o sistemes de recomanació (Ingram and Johnson 2022). Permet recuperar grans quantitats de dades (Reutter and Spiíker 2019, Jaillant, Aske, and Caputo 2024), reduint la càrrega manual d'aquests processos (Venkata et al. 2022), així com les seves deficiències (Omigie, Krubu, and Anthony 2023).

En relació amb la gestió, l'aplicació de la IA ajuda en la realització d'aquesta tasca, millorant la precisió, la confiabilitat i la innovació (Omigie, Krubu, and Anthony 2023), i permet alleugerir tots aquells problemes que puguin derivar-se'n (Modiba, Ngoepe, and Ngulube 2019). Permet modernitzar i millorar els sistemes de gestió documental (Stylianou et al. 2022, Omigie, Krubu, and Anthony 2023) creant un entorn molt més eficaç. Pot promoure la millora i creació de noves eines adaptades a l'àmbit arxivístic (Duranti 2022), ajudant a superar les limitacions dels arxius tradicionals (Colavizza et al. 2022), i transformant el seu paper en la gestió de les dades (Reutter and Spiíker 2019). En conseqüència, es promouen noves formes d'arxiu més democràtiques i inclusives en la recopilació i anàlisi de dades (Chabin 2020, Colavizza et al. 2022).

Relacionat amb la seguretat i protecció de les dades, l'aprenentatge automàtic permet desenvolupar eines que poden identificar automàticament arxius confidencials (Baron and Payne 2017, Jaillant 2022b) amb informació sensible i protegida i, per tant, catalogar, codificar i indexar de manera segura (Omigie, Krubu, and Anthony 2023). En conseqüència, permet controlar l'accés no autoritzat (Modiba 2022b) i prevenir els perills relacionats amb la seva protecció (Omigie, Krubu, and Anthony 2023).

En cinquè lloc, l'ús de la IA millora la qualitat i facilita l'accés a grans volums d'informació (Gelashvili and Pappel 2021). Permet gestionar els permisos d'accés basant-se en rols, comportaments i necessitats (Baron and Payne 2017), resolent gran part dels problemes d'accés generats (Jaillant, Aske, and Caputo 2024).

L'aplicació de sistemes d'automatització, tant en la presa de decisions com en els fluxos de treball, permeten optimitzar la gestió documental (Alaoui 2023) amb una major transparència i responsabilitat (Gelashvili and Pappel 2021). Un fet que deriva en processos més ràpids i amb menys pèrdua de dades derivades dels errors humans

(Azzam et al. 2023), juntament amb la possibilitat de filtrar i processar grans quantitats de dades (Venkata et al. 2022), amb una millora constant de la seva capacitat d'aprenentatge (Venkata, Young, and Green 2021).

En relació amb la conservació i manteniment, els beneficis són principalment aquells relacionats amb una major estabilitat i rapidesa dels sistemes d'emmagatzematge i preservació de la documentació (Omigie, Krubu, and Anthony 2023, Jaillant, Aske, and Caputo 2024), així com fomentar processos de digitalització més àgils i eficaços (T. Modiba et al., 2019).

L'ús de la IA envers el compliment normatiu, permet assegurar l'aplicació de les lleis i regulacions de manera actualitzada (Chun 2007). Promou una major garantia en els controls de qualitat, amb la possibilitat d'auditar de manera automàtica i eficient, i gestionar sistemàticament els processos de valoració de la documentació (Omigie, Krubu, and Anthony 2023), així com els processos de rendició de comptes (Stylianou et al. 2022). Permet gràcies als seus models, comprovar la veracitat de la informació (Wang 2022) i garantir l'autenticitat dels documents (Jaillant 2022b).

En l'aspecte econòmic, ofereix una millor rendibilitat operativa (Omigie, Krubu, and Anthony 2023) gràcies a la seva capacitat d'aprenentatge continu, establint així una bona relació cost – eficàcia (Gelashvili and Pappel 2021) i fomentant la reducció de costos (Azzam et al. 2023). A més, té la capacitat de generar més oportunitats laborals per als joves, disminuint els problemes de desocupació gràcies a la creació i modificació de nous llocs de treball (Omigie, Krubu, and Anthony 2023). Com a resultat, facilita el creixement econòmic del sector públic (Chun 2007).

En últim lloc, i en relació amb el servei, l'ús d'IA permet millorar el temps de resposta (Baron and Payne 2017), augmentant la qualitat i productivitat gràcies a la dinamització dels processos i fluxos de treball (Chun 2007). Ajuda a minimitzar l'error humà, ja que és possible dur a terme tasques monòtones amb més exactitud i precisió (Omigie, Krubu, and Anthony 2023), i per tant oferir millors serveis i més personalitzats (Reutter and Spíker 2019). En relació amb els/les usuaris/es, això els proporciona una millor experiència (Ingram and Johnson 2022), que enriqueix la relació entre l'administració i la ciutadania (Chun 2007) mitjançant una interacció més directa (Omigie, Krubu, and Anthony 2023). D'altra banda, l'automatització permet reduir la càrrega de treball dels empleats públics, deixant-los més temps per centrar-se en altres aspectes de la seva feina (Stylianou et al. 2022).

A banda dels beneficis, també s'han identificat els riscos de l'aplicació de la IA als arxius de les administracions públiques. Novament, per una millor comprensió, els riscos s'han agrupat en 8 categories (**Taula 9**).

Categoria	Riscos
1) Privacitat i seguretat de les dades	Riscos en la privacitat i seguretat de les dades (Reutter and Spiiker 2019) Problemes de seguretat en relació amb dades sensibles (Venkata et al. 2022) La manca de control pot posar en risc dades sensibles (Jaillant and Rees 2023) Baixa precisió en la verificació de dades (Wang 2022) L'accés als "arxius foscos" pot provocar la divulgació errònia d'informació confidencial (Jaillant 2022b)
2) Biaixos i discriminació	Biaixos en els models algorítmics (The National Archives 2021) Possibles biaixos en la presa de decisions (Venkata et al. 2022) Biaixos i conseqüències ètiques (Colavizza et al. 2022) Biaixos (Duranti 2022) Biaixos actuals (Ingram and Johnson 2022) Risc d'ampliar el biaix algorítmic, reforçant estereotips i percepcions esbiaixades (Jaillant, Aske, and Caputo 2024) Biaixos en grans conjunts de dades (Jaillant 2022b) Les conseqüències dels biaixos algorítmics (sexisme, discriminació...) (Jaillant and Rees 2023) Proporcionen evidència basada en probabilitats, no en fets (Duranti 2022) Depenen directament de les dades proporcionades per les persones (Duranti 2022) No pot discernir el que és moralment o socialment acceptat del que no ho és (Duranti 2022) Augment dels prejudicis falta d'ètica (Jaillant, Aske, and Caputo 2024) Riscos ètics (Jaillant 2022b) Manca d'ètica en la presa de decisions (Jaillant, Aske, and Caputo 2024) Aspectes ètics (Lindgren 2022) Manca d'ètica (Reutter and Spiiker 2019)
3) Dependència i confiabilitat	Manca de confiabilitat i equitat (Jaillant, Aske, and Caputo 2024) Manca de confiança (Colavizza et al. 2022) No pot reemplaçar algunes de les tasques realitzades pels arxivers (The National Archives 2021) Impossibilitat de tractar alguns documents de text escrits o anotats a mà mitjançant OCR (Ingram and Johnson 2022) Falta de pensament propi i creativitat (Omigie, Krubu, and Anthony 2023)
4) Transparència i explicabilitat (XAI)	Manca de responsabilitat i transparència (Jaillant, Aske, and Caputo 2024) Manca de transparència dels sistemes d'IA (Lindgren 2022) La manca de transparència constitueix una greu amenaça democràtica (Reutter and Spiiker 2019) Falta de transparència i explicabilitat (XAI) (Venkata et al. 2022) Opacitat dels algorismes d'IA (Rolan et al. 2019) Manca de transparència i rendició de comptes en els processos de presa de decisions automatitzada (Jaillant and Rees 2023) Proporcionen proves que no són interpretables ni transparents (Duranti 2022) Funcionament massa opac per complir amb els estàndards de responsabilitat i transparència (Davet, Hamidzadeh, and Franks 2023) Falta de coneixement sobre les conseqüències (Colavizza et al. 2022) Impossibilitat de documentar la racionalitat que hi ha darrere les decisions de la IA (Lindgren 2022)
5) Compliment normatiu i legal	Manca de regulació sobre les actuacions de la IA (Lindgren 2022) Manca de legislació en alguns països (p.e. Sudàfrica) (Modiba 2022b) Creació de problemes informatius, ètics i legals a les organitzacions (Alaoui 2023)
6) Qualitat de les dades	Dades i algorismes parcials o dubtosos (Omigie, Krubu, and Anthony 2023) L'exactitud i precisió de les dades depenen estrictament dels conjunts de dades disponibles en l'aprenentatge de l'algorisme (Venkata, Young, and Green 2021) La presa de decisions equívokes pot comportar pèrdua d'informació, contingències textuais i llacunes interpretatives (Reutter and Spiiker 2019) El model resultant depèn directament de la precisió en l'entrenament de l'algorisme (Rolan et al. 2019)

7) Impacte laboral i econòmic	Augment de les responsabilitats en l'àmbit laboral (Gelashvili and Pappel 2021) Els nous requisits de documentació de la Llei d'IA augmenten la càrrega administrativa (Lindgren 2022) Els resultats poden desafiar l'autonomia i privacitat de les persones (Duranti 2022) Por de perdre la feina (Gelashvili and Pappel 2021) Creació de desocupació en certs sectors (Omigie, Krubu, and Anthony 2023) Augment de la despesa econòmica (Gelashvili and Pappel 2021) Alts costos de desenvolupament (Lindgren 2022) Alts costos d'instal·lació (Omigie, Krubu, and Anthony 2023)
8) Integritat de les dades	Riscos relacionats amb l'emmagatzematge de documents i la conservació de les tasques realitzades (Gelashvili and Pappel 2021) Corrupció d'arxius o metadades a causa de la migració (ML) (Venkata et al. 2022) Falta de respecte als principis ètics arxivístics (Davet, Hamidzadeh, and Franks 2023)

Taula 9. Riscos de l'ús de la IA en els arxius públics a partir dels resultats dels estudis seleccionats.

En primer lloc, els riscos relacionats amb la privacitat i seguretat de les dades derivats de l'ús de la IA en els arxius de les administracions públiques, se centren en els problemes al voltant de les dades sensibles (Venkata et al. 2022), com la manca de control (Jaillant and Rees 2023) o la baixa precisió en la verificació de les dades (Wang 2022). L'accés no controlat als anomenats "arxius foscos" pot provocar la divulgació errònia d'informació confidencial (Jaillant 2022b).

En segon lloc, un dels riscos més percebuts són aquells al voltant dels biaixos generats pels models algorítmics en la presa de decisions (The National Archives 2021). Un fet que pot generar estereotips i prejudicis que provoquin conseqüències com el sexisme, la discriminació (Jaillant and Rees 2023) i en definitiva, una manca d'ètica en les decisions (Reutter and Spiíker 2019), degut a la impossibilitat de poder discernir el que és moral del que no (Duranti 2022).

En relació amb la dependència i confiabilitat, els riscos més habituals són la manca de confiança en aquestes noves eines (Colavizza et al. 2022). En molts casos la IA no és capaç de realitzar algunes de les tasques pròpies dels/les arxiviers/es (The National Archives 2021), com és el cas del tractament amb OCR d'alguns documents de text escrits o anotats a mà (Ingram and Johnson 2022). L'abús i dependència d'aquestes eines, però també genera altres inquietuds com són la falta de pensament propi i la creativitat (Omigie, Krubu, and Anthony 2023).

En quart lloc, un dels riscos més habituals generat entorn la IA és la manca de transparència i responsabilitat (Jaillant, Aske, and Caputo 2024) en els processos de presa de decisions automatitzada i sobre les conseqüències del seu ús (Colavizza et al. 2022), una opacitat que en el cas de les administracions públiques constitueix una greu amenaça democràtica (Reutter and Spiíker 2019). La manca d'explicabilitat (XAI)

(Venkata et al. 2022) o la impossibilitat de documentar la racionalitat que hi ha darrere les decisions preses per IA, també suposa un risc a tenir en compte (Lindgren 2022).

Sobre el compliment normatiu i legal, les principals preocupacions són aquelles relacionades amb una manca de regulació sobre les actuacions dutes a terme amb IA (Lindgren 2022), un fet que pot provocar problemes informatius, ètics i legals a les organitzacions (Alaoui 2023).

La qualitat de les dades en l'ús de la IA poden ser parcials o dubtoses (Omigie, Krubu, and Anthony 2023), ja que depenen estrictament de l'exactitud i precisió dels conjunts de dades disponibles en l'aprenentatge de l'algoritme (Venkata, Young, and Green 2021). En conseqüència, la presa de decisions equívokes pot provocar la pèrdua d'informació, així com llacunes interpretatives (Reutter and Spíker 2019).

En relació amb l'impacte laboral, els principals riscos descrits són aquells que es relacionen amb un augment de les responsabilitats (Gelashvili and Pappel 2021), i que impliquen un augment de la càrrega administrativa (Lindgren 2022). Altres temors són els relacionats amb la por de perdre la feina (Gelashvili and Pappel 2021) i a la creació de desocupació en certs sectors (Omigie, Krubu, and Anthony 2023). En el cas dels riscos econòmics, els alts costos de desenvolupament i instal·lació (Lindgren 2022, Omigie, Krubu, and Anthony 2023) provoquen un augment de la despesa econòmica de les organitzacions (Gelashvili and Pappel 2021).

Finalment, la integritat de les dades és un dels aspectes més compromesos en l'ús de la IA. La corrupció d'arxius o metadades causades per la migració en l'aplicació de l'aprenentatge automàtic (Venkata et al. 2022), o els riscos relacionats amb l'emmagatzematge i la conservació (Gelashvili and Pappel 2021) poden provocar una falta de respecte als principis ètics arxivístics (Davet, Hamidzadeh, and Franks 2023), entre els quals la integritat dels documents resulta un dels pilars fonamentals.

4.4.2 Requisits en l'aplicació de la IA als arxius de les administracions públiques

A banda dels beneficis i riscos, en la mostra d'estudis seleccionada s'han detectat els requisits necessaris per mantenir i desenvolupar correctament i de manera sostenible l'aplicació de la IA. A continuació es presenta en la **Taula 10** una descripció general d'aquests, organitzats en 7 categories.

Categoria	Requisits
1) Dades i infraestructura	Necessitat de grans conjunts de dades, amb les que els algoritmes puguin provar-se i perfeccionar-se (Reutter and Spíker 2019)

	Necessitat d'un gran nombre de dades pel correcte funcionament del model (Venkata et al. 2022) La qualitat de les dades d'entrenament és fonamental per obtenir bons resultats (Venkata et al. 2022) Importància de la identificació i preparació de les dades d'entrenament i prova (Rolan et al. 2019) Invertir en la preparació de les dades per als processos de prova i perfeccionament dels models (The National Archives 2021) Capturar i preservar informació sobre la IA com a agent (Davet, Hamidzadeh, and Franks 2023) Avaluar, actualitzar i entrenar els models algorítmics contínuament a mesura que s'obtinguin noves dades, enfocaments i tècniques (Ingram and Johnson 2022) Revisar periòdicament els models (Venkata, Young, and Green 2021) Mantenir actualitzat el mòdul d'IA després de la seva implementació, per reflectir els canvis legislatius i operatius de l'organització (Chun 2007) La participació dels arxivers en el disseny de la infraestructura és fonamental per garantir que es compleixen les necessitats dels arxius (Lindgren 2022) Tenir en compte el paper dels arxius en la creació de conjunts de dades per l'aprenentatge dels algoritmes (Jaillant, Aske, and Caputo 2024) Crear eines dissenyades específicament per arxius (Duranti 2022)
2) Organització	Transmetre la necessitat i l'efecte positiu del canvi mitjançant la implicació de la direcció de l'organització (Gelashvili and Pappel 2021) Fomentar la interoperabilitat i la normalització de sistemes (Gelashvili and Pappel 2021) És necessari un major desenvolupament de la tecnologia (Colavizza et al. 2022) Tenir en compte factors propis i requisits específics a l'hora d'escollir una eina (The National Archives 2021)
3) Capacitació i formació	Desenvolupar projectes conjunts entre investigadors en ciències de la informació i empreses de tecnologia proveïdores d'IA (Chabin 2020) Col·laboració amb proveïdors d'eines d'IA i influir en el seu desenvolupament futur (The National Archives 2021) Col·laboració amb els proveïdors d'eines d'IA (Venkata et al. 2022) Més investigació i treball directe amb els arxius (Colavizza et al. 2022) Treballar en col·laboració mitjançant la creació d'equips interdisciplinaris (Jaillant 2022b) Estrènyer la col·laboració entre professionals relacionats amb la gestió documental (Jaillant and Rees 2023) Fomentar l'acció col·laborativa d'estratègia digital per assegurar l'accés futur i l'ús dels arxius digitals (Jaillant, Aske, and Caputo 2024) Millorar la comprensió dels conceptes d'IA entre els arxivers (Rolan et al. 2019) Capacitar els equips de gestió documental en coneixements sobre dades i IA, i estrènyer les relacions amb experts en el sector (Venkata et al. 2022) Formar tècnicament als arxivers i potenciar les relacions amb professionals en anàlisi de dades (The National Archives 2021) Formació adequada en TI (Gelashvili and Pappel 2021) Capacitar al personal en habilitats digitals per evitar que prefereixin realitzar tasques manuals que els proporcionen seguretat, però més temps de dedicació (Gelashvili and Pappel 2021) Capacitar al personal en l'ús de les noves tecnologies i mantenir-lo actualitzat constantment (Ingram and Johnson 2022) Capacitar al personal encarregat sobre les pautes, regulacions i bones pràctiques per gestionar els riscos de la IA (Ingram and Johnson 2022) Desenvolupar una comprensió per part dels arxivers envers les noves tecnologies amb IA (Lindgren 2022) Participació activa dels arxivers en els debats sobre IA aplicada als arxius (Jaillant 2022b) Desenvolupar més estudis de cas que facilitin la seva aplicació als arxius (Lindgren 2022) Actualitzar els fonaments teòrics de la pràctica arxivística amb les recents (Colavizza et al. 2022) Formar sobre els programes i l'aplicació de la IA en la gestió dels arxius (Modiba 2022b)
4) Ètics	Transparència (Jaillant and Rees 2023) Crear conjunts de dades sobre els conceptes de consentiment, inclusió, poder, transparència, ètica i privacitat (Jaillant, Aske, and Caputo 2024) Garantir la transparència dels sistemes d'IA (Lindgren 2022) Tractar els biaixos associats als algoritmes de "caixa negra" (Jaillant 2022b) Evitar la introducció d'errors o biaixos en l'entrenament algorítmic (Rolan et al. 2019)

	Publicar els processos i la lògica subjacent a la gestió i preservació de les dades d'entrenament per identificar i evitar possibles biaixos (Venkata et al. 2022) Representar les dades de forma equilibrada per evitar biaixos en la presa de decisions (Venkata et al. 2022) Supervisió adequada per evitar problemes informatius, ètics i legals (Alaoui 2023) El manteniment dels arxius pot contribuir al desenvolupament i ús adequat de la IA en relació amb la procedència, avaluació, transparència i responsabilitat (Colavizza et al. 2022) Els arxiviers han de considerar les seves estratègies per mantenir la responsabilitat dels arxius en l'era de la IA (Davet, Hamidzadeh, and Franks 2023) Desenvolupar programes de capacitació ètica i comprensió conscient dels riscos, resultats i conseqüències de la IA (Ingram and Johnson 2022) Millorar la confiança en les tècniques d'IA mitjançant el desenvolupament d'un marc ètic més sòlid i una millor comprensió del seu impacte (Colavizza et al. 2022) La col·laboració entre arxiviers i investigadors, amb una ètica professional compartida, ajudaria a generar més confiança en les eines d'IA (Jaillant and Rees 2023)
5) Econòmics	Més atenció als mètodes i recursos necessaris per produir resultats òptims (Reutter and Spíker 2019) Dotar dels recursos adequats (Rolan et al. 2019) Gran quantitat de recursos (Venkata et al. 2022) Equipar el personal de gestió documental amb les eines adequades per la mineria i l'anàlisi de dades (Venkata et al. 2022) Emprendre projectes propis (ML) que permetin prescindir de les eines propietàries dels gegants tecnològics (Jaillant 2022b)
6) Legals	Les organitzacions són les encarregades d'aplicar la legislació pertinent, i assegurar que aquesta es compleixi adequadament (Modiba 2022b) Creació de normes aplicades a tots els arxius i a altres àmbits (Davet, Hamidzadeh, and Franks 2023) Complir amb les regulacions de protecció de dades (Jaillant 2022b) Proporcionar sistemes d'auditoria continus mitjançant la implementació de noves estructures organitzatives (Ingram and Johnson 2022) Potenciar que les institucions registrin models i algoritmes que puguin ser auditable (Jaillant 2022b) Rendició de comptes (Jaillant and Rees 2023)
7) Socials	Potenciar la creació d'una IA explicable (XAI) que tingui en compte les persones i l'entorn en què operarà (Jaillant, Aske, and Caputo 2024)

Taula 10. Requisits en l'ús de la IA als arxius públics, a partir dels resultats dels estudis seleccionats.

En primer lloc, un dels requisits principals relacionats amb les dades és la necessitat de grans volums que permetin als models algorítmics provar-se i perfeccionar-se progressivament (Reutter and Spíker 2019), així com la importància de la qualitat de les dades per obtenir bons resultats (Venkata et al., 2022). Invertir en la identificació i preparació de les dades (The National Archives 2021), així com capturar i preservar la informació de la IA com a agent (Davet, Hamidzadeh, and Franks 2023), permetrà crear models més sòlids i precisos. Pel que fa a la infraestructura, és necessària una revisió periòdica a mesura que s'obtenen noves dades, enfocaments i tècniques (Ingram and Johnson 2022), mantenint-la actualitzada davant els canvis legislatius i operatius de l'organització (Chun 2007). A més, és important tenir en compte la participació dels/les arxiviers/es en el seu disseny per garantir el correcte compliment dels requisits específics del sector (Lindgren 2022).

En relació amb els requisits organitzatius, és fonamental la implicació de la direcció, amb l'objectiu de transmetre la necessitat i efecte positiu del canvi (Gelashvili and Pappel

2021). De la mateixa manera, és important fomentar la interoperabilitat entre organitzacions i la normalització de sistemes (Gelashvili and Pappel 2021), tot i que encara és necessari un major desenvolupament de les tecnologies amb IA (Colavizza et al. 2022). Tot i això, és fonamental tenir en compte factors propis i requisits específics a l'hora d'escollir una eina tecnològica (The National Archives 2021), ja que cada organització té necessitats úniques que han de ser considerades per tal d'assegurar que les solucions s'adaptin als seus objectius.

En tercer lloc, la capacitat dels/les arxivers/es en coneixements sobre noves tecnologies vinculades a la IA (Gelashvili and Pappel 2021) i la seva formació contínua (Ingram and Johnson 2022), és un requisit indispensable per abordar els reptes que presenten els arxius digitals al voltant d'aquestes noves eines. Millorar la comprensió d'aquests nous conceptes (Rolan et al. 2019), juntament amb la creació d'equips interdisciplinaris amb professionals en anàlisi de dades (Jaillant 2022b), o fins i tot, projectes conjunts entre investigadors/es i empreses de tecnologia proveïdores d'IA (Chabin 2020), permetran assegurar l'accés futur i l'ús dels arxius digitals (Jaillant, Aske, and Caputo 2024). Finalment, és important reivindicar la participació activa dels/les arxivers/es en els debats sobre IA aplicada als arxius (Jaillant 2022b), així com el desenvolupament de més estudis de cas que en facilitin l'aplicació (Lindgren 2022).

En el camp de la IA, l'ètica és un aspecte fonamental a tenir en compte. Garantir la transparència dels sistemes d'IA (Lindgren 2022) és essencial per prevenir errors o biaixos durant l'entrenament dels algorismes (Rolan et al. 2019). La publicació dels processos i la lògica subjacents en la gestió i preservació de les dades d'entrenament, així com la representació equilibrada de les dades, pot minimitzar aquests biaixos en la presa de decisions (Venkata et al. 2022). Tanmateix, una supervisió adequada i el manteniment dels arxius poden contribuir al desenvolupament i ús correcte de la IA, evitant problemes informatius, ètics i legals (Alaoui 2023). El desenvolupament de programes de capacitat ètica (Ingram and Johnson 2022) i la col·laboració entre arxivers/es i investigadors/es, amb una ètica professional compartida, poden ajudar a comprendre els riscos i conseqüències de la IA, generant més confiança en aquestes tecnologies (Jaillant and Rees 2023).

En l'aspecte econòmic i per aconseguir resultats òptims en la gestió documental, és fonamental prestar atenció als mètodes i recursos necessaris (Reutter and Spíker 2019). Això implica no només dotar dels recursos adequats (Rolan et al. 2019), sinó també garantir una gran quantitat de recursos disponibles per afrontar els desafiaments que es puguin presentar (Venkata et al. 2022). Així mateix, és essencial dotar el personal amb les eines adequades per la mineria i anàlisi de dades (Venkata et al. 2022), un fet

que permetrà una gestió més eficient de la informació. A més, emprendre i invertir en projectes propis d'aprenentatge automàtic pot ser una estratègia clau per prescindir de les eines propietàries dels gegants tecnològics, oferint una major autonomia i flexibilitat en la gestió de les dades (Jaillant 2022b).

La creació de normes en l'àmbit legal aplicades a tots els arxius i a altres àmbits és essencial per garantir una gestió eficient i segura de la informació (Davet, Hamidzadeh, and Franks 2023). Les organitzacions són responsables d'aplicar la legislació pertinent, i assegurar que aquesta es compleixi adequadament (Modiba 2022b). De la mateixa manera succeeix amb altres regulacions més específiques, com la de protecció de dades (Jaillant 2022b), que requereix una atenció constant per adaptar-se als constants canvis legislatius i tecnològics. Per assegurar el correcte compliment d'aquestes normes, és important proporcionar sistemes d'auditoria continus mitjançant la implementació de noves estructures organitzatives (Ingram and Johnson 2022), un fet que resulta més fàcil si des d'un inici es potencia que les institucions registrin models i algoritmes que puguin ser auditables fàcilment (Jaillant 2022b). Finalment, la rendició de comptes (Jaillant and Rees 2023) és un altre aspecte fonamental en la gestió de la informació i les dades, ja que permet a les organitzacions ser transparents i responsables davant el compliment de les seves accions.

En darrer lloc, un dels requisits socials en l'ús de la IA recau principalment en potenciar la creació d'una IA explicable (XAI), no només des d'una perspectiva tecnològica i algorítmica, sinó que tingui en compte les persones i l'entorn social en què operarà (Jaillant, Aske, and Caputo 2024). Un enfocament que busca no només avançar en la capacitat tècnica de les IA sinó també assegurar que aquestes tecnologies siguin comprensibles i transparents per a la societat.

4.4.3 Reptes de futur pel desenvolupament i aplicació de la IA als arxius de les administracions públiques

Un altre aspecte a tenir en compte, segons els resultats obtinguts de la mostra d'estudis seleccionats, són els reptes de futur. Aquests, fan referència a les tasques o investigacions necessàries per continuar desenvolupant l'aplicació de la IA en l'àmbit dels arxius de les administracions públiques. Com s'ha fet en els casos anteriors, i per facilitar la comprensió, els resultats s'han organitzat mitjançant 7 categories (**Taula 11**).

Categoria	Reptes de futur
1) Aplicació i millora	Aplicar nous sistemes d'IA en més contextos reals (Azzam et al. 2023) Implementar de forma pràctica en un entorn real (Hu et al. 2005) Posar a prova de forma pràctica els models creats (Gelashvili and Pappel 2021)

	Avaluar la creació i aplicació pràctica de diferents tipus d'aprenentatge automàtic (Rolan et al. 2019) Continuar investigant sobre l'aplicació de la IA més enllà de les fronteres disciplinàries pròpies dels arxius (Jaillant and Rees 2023) Crear equips diversos per treballar en el disseny i la implementació de sistemes (Ingram & Johnson, 2022) Millorar els models creats (Azzam et al. 2023) Afegir noves característiques per entrenar i millorar els models (Venkata, Young, and Green 2021) Contribuir a l'eficiència en la gestió de la informació governamental (Lindgren 2022) Invertir en l'adopció i aplicació de tecnologies d'IA en el sector públic de països com Sud-àfrica (Modiba, Ngoepe, and Ngulube 2019) Potenciar les eines d'IA a les organitzacions públiques nigerianes (Omigie, Krubu, and Anthony 2023) Investigar, desenvolupar i comparar models més efectius i eficients (Hu et al. 2005) Experimentar de forma pràctica amb els productes d'IA disponibles al mercat (Venkata et al. 2022) Analitzar l'equilibri entre la facilitat d'ús i les opcions de configuració dels productes d'IA disponibles (Venkata et al. 2022)
2) Anàlisi de continguts i metadades	Millorar la tècnica de classificació de documents amb metadades utilitzant tècniques d'aprenentatge profund (Alothman and Sait 2022) Investigar sobre la classificació de documents, combinant l'aprenentatge automàtic i el coneixement humà (Venkata et al. 2022) Desenvolupar enfocaments basats en xarxes neuronals (DL) per la classificació de nous formats d'arxiu amb característiques similars (Venkata, Young, and Green 2021) Realitzar nous experiments per ampliar les capacitats d'anàlisi de l'eina (Venkata, Young, and Green 2021)
3) Accés i recuperació	Avaluar les necessitats d'accés dels usuaris (Hu et al. 2005)
4) Privacitat i seguretat	Utilitzar la IA basada en conceptes d'arxiu per autenticar documents i detectar qualsevol interferència entre ells (Duranti 2022) Continuar investigant sobre la revisió automàtica aplicada a la sensibilitat i confidencialitat de les dades (Jaillant 2022b) Contribuir al debat sobre arxius, transparència i IA examinant els models creats (Lindgren 2022)
5) Automatització de processos	Avançar en els processos automatitzats de presa de decisions (Gelashvili and Pappel 2021) Continuar desenvolupant mètodes d'aprenentatge conjunt per permetre que els sistemes automatitzin la selecció d'algoritmes i l'ajust dels paràmetres (Baron and Payne 2017) Investigar sobre mètodes eficaços per seleccionar automàticament un nombre més gran de paraules durant el procés de mostreig i avaluar el seu impacte (Stylianou et al. 2022) Caracteritzar de forma teòrica i pràctica la interacció entre la recopilació de paradades i els seus efectes en l'automatització arxivística basada en IA (Davet, Hamidzadeh, and Franks 2023)
6) Humanitats digitals	Mesurar l'impacte dels sistemes creats en la confiança de les persones en casos d'ús específics (Azzam et al. 2023) Ampliar els estudis d'usuaris i explorar els factors humans que juguen un paper fonamental en l'acceptació dels serveis amb IA (Gelashvili & Pappel, 2021) Investigar i verificar si l'actitud positiva del sector públic cap a la IA és realment tan comuna (Lindgren 2022) Investigar com la IA modifica el rol professional dels arxivistes (Lindgren 2022) Conèixer com afectaran els recursos humans, tecnològics i financers a la capacitat dels arxivistes per recopilar i utilitzar les paradades (Davet, Hamidzadeh, and Franks 2023)

7) Aspectes ètics i legals	Elaborar estratègies eficients per crear normes i directrius sobre paradades en l'àmbit dels arxius (Davet, Hamidzadeh, and Franks 2023) Explorar com s'aplicaran els principis de la norma ISO 24143 (governança de la informació) tenint en compte els requisits de la IA i en concordança amb la cultura organitzativa (Alaoui 2023) Dissenyar i implementar sistemes d'IA alineats èticament (Ingram and Johnson 2022) Adoptar un enfocament sistemàtic per auditar els efectes del biaix algorítmic en la IA (Ingram & Johnson, 2022)
----------------------------	---

Taula 11. Reptes de futur en l'ús de la IA als arxius públics, a partir dels resultats dels estudis seleccionats.

L'aplicació pràctica de sistemes d'IA en més quantitat de contextos reals, és fonamental per a maximitzar els seus beneficis en el futur (Azzam et al. 2023). És necessari continuar investigant més enllà dels límits disciplinaris tradicionals dels arxius (Jaillant and Rees 2023), i formar equips diversos per al disseny i la implementació d'aquests sistemes (Ingram & Johnson, 2022). Aquest enfocament contribuirà a millorar els models, incorporant noves característiques d'entrenament i millora continua (Venkata, Young, and Green 2021). Contribuir a l'eficiència en la gestió de la informació governamental (Lindgren 2022) pot tenir un impacte significatiu en la presa de decisions i la transparència de les organitzacions públiques. Per això, països com Sud-àfrica i Nigèria tenen el repte d'adoptar i aplicar tecnologies d'IA en el seu sector públic (Modiba, Ngoepe, and Ngulube 2019, Omigie, Krubu, and Anthony 2023). Per tant, és essencial prioritzar la investigació, el desenvolupament i comparació de models cada vegada més efectius i eficients (Hu et al., 2005), així com experimentar de manera pràctica amb productes ja disponibles (Venkata et al. 2022), per garantir una execució correcta i un impacte positiu de la IA.

Els reptes futurs en l'anàlisi de continguts i metadades als estudis seleccionats se centren principalment en la investigació i millora de les tècniques de classificació automàtica de documents. Això implica la combinació de tècniques d'aprenentatge profund, com les xarxes neuronals (Alothman and Sait 2022, Venkata, Young, and Green 2021) i d'aprenentatge automàtic (Venkata et al. 2022). En conseqüència, és fonamental dur a terme nous experiments per ampliar les capacitats d'anàlisi de les eines (Venkata, Young, and Green 2021), amb l'objectiu de millorar els processos de classificació, fent-los més precisos i adaptats a les necessitats específiques de cada context.

En el context de l'accés i la recuperació de la informació, el repte identificat en el conjunt de la mostra ha estat la de dur a terme una avaluació formal de les necessitats d'accés dels/les usuaris/es. Una tasca fonamental, ja que un dels objectius finals de l'estudi era

atendre millor les necessitats d'informació dels/les usuaris/es en una organització governamental (Hu et al. 2005).

En relació amb la privacitat i seguretat de les dades, és necessari promoure l'ús de la IA basada en conceptes d'arxiu per autenticar documents i detectar qualsevol interferència (Duranti 2022). Tanmateix, és important continuar investigant sobre la revisió automàtica aplicada a les dades sensibles i confidencials (Jaillant 2022b), contribuint a millorar la protecció de la privacitat i seguretat de la informació. Per a això, cal examinar des d'una perspectiva arxivística els models ja existents, amb l'objectiu d'enriquir el debat sobre arxius, transparència i IA (Lindgren 2022).

En l'automatització de processos, avançar en la presa de decisions (Gelashvili and Pappel 2021) i continuar desenvolupant mètodes d'aprenentatge conjunt (Baron and Payne 2017), millorarà l'adaptació dels sistemes a les necessitats dels/les usuaris/es i dels entorns. Algunes de les aplicacions destacades en els estudis seleccionats inclouen l'automatització de la selecció d'algoritmes i l'ajust de paràmetres, així com la selecció automàtica d'un major nombre de paraules durant el mostreig i l'avaluació del seu impacte (Stylianou et al. 2022). Aquestes són tasques essencials per millorar la representació i generalització dels models. A més, caracteritzar de manera teòrica i pràctica la interacció entre la recopilació de "paradades" i els seus efectes en l'automatització arxivística basada en IA (Davet, Hamidzadeh, and Franks 2023), pot proporcionar una comprensió més profunda dels processos arxivístics i permetrà una millor implementació de la IA.

En l'àmbit de les humanitats digitals i la IA als arxius, un dels reptes de futur és mesurar l'impacte d'aquests nous sistemes en la confiança de les persones i en casos d'ús específics (Azzam et al. 2023), per assegurar una adopció satisfactòria de la tecnologia. Ampliar els estudis d'usuaris/es i explorar els factors humans que contribueixen en l'acceptació de serveis amb IA (Gelashvili & Pappel, 2021), permetrà adaptar les solucions tecnològiques a les necessitats reals. Per aquest motiu, és interessant investigar i verificar si l'actitud positiva del sector públic és realment tan comuna com indiquen alguns estudis (Lindgren 2022). Investigar com la IA modifica el rol professional dels/les arxivers/es i conèixer com afectaran els recursos humans, tecnològics i financers a la seva capacitat per recopilar i utilitzar les "paradades" (Davet, Hamidzadeh, and Franks 2023), proporcionarà una visió integral sobre tot el que la IA ofereix al sector dels arxius.

Finalment, en el context dels aspectes ètics i legals, alguns dels reptes de futur identificats en la mostra són sobre casos concrets, com el desenvolupament

d'estratègies eficients per establir normes i directrius sobre l'ús de “paradades” als arxius (Davet, Hamidzadeh, and Franks 2023), o sobre com explorar l'aplicació dels principis de la norma ISO 24143 sobre governança de la informació, tenint en compte els requisits de la IA i en concordança amb la cultura organitzativa de l'entitat (Alaoui 2023). Així mateix, dissenyar i implementar sistemes d'IA alineats èticament és fonamental per garantir el respecte als valors ètics i als drets dels/les usuaris/es. Per això, adoptar un enfocament sistemàtic per auditar els efectes del biaix algorítmic de la IA (Ingram & Johnson, 2022) pot mitigar els riscos associats a la discriminació en el seu ús.

4.4.4 Abast de la recerca sobre l'aplicació de la IA als arxius de les administracions públiques

En aquesta secció es descriu l'abast utilitzat en els estudis seleccionats, incloent-hi el tipus d'IA, l'àmbit arxivístic d'aplicació i el tipus d'organització. En primer lloc, s'examina el tipus d'IA en què se centren els estudis i s'observa que una gran part ($n = 11$) utilitza el terme “IA” de manera genèrica, sense especificar-ne el tipus. Tot i això, en quatre casos, es fa referència a l'automatització de processos intel·ligents (IPA) com a forma d'aplicació d'aquesta IA, mencionada de forma àmplia i poc específica. La major part dels estudis es refereixen a un tipus particular d'IA, com és l'aprenentatge automàtic (*Machine Learning – ML*) ($n=18$), l'aprenentatge profund (*Deep Learning – DL*) ($n=7$) o els sistemes experts ($n=1$). És important assenyalar que la major part dels estudis no se centren en un únic tipus, sinó que combinen l'anàlisi o l'ús de diversos tipus o tècniques. Així, dins l'aprenentatge automàtic (ML) trobem combinacions que inclouen el processament de llenguatge natural (NLP) ($n=8$), els sistemes de recomanació ($n=1$) i el reconeixement d'entitats amb nom (NER) ($n=1$). En el cas de l'aprenentatge profund (DL), un estudi combina aquesta tècnica amb xarxes neuronals recurrents (RNN). En l'únic cas documentat d'IA basat en sistemes experts (Chun 2007), es fa referència al raonament basat en casos (CBR). Finalment, una altra tècnica àmpliament documentada és el reconeixement òptic de caràcters (OCR) ($n=6$) que, tot i no ser un tipus d'IA, hi està molt vinculada, especialment en l'àmbit dels arxius (**Fig. 8**).

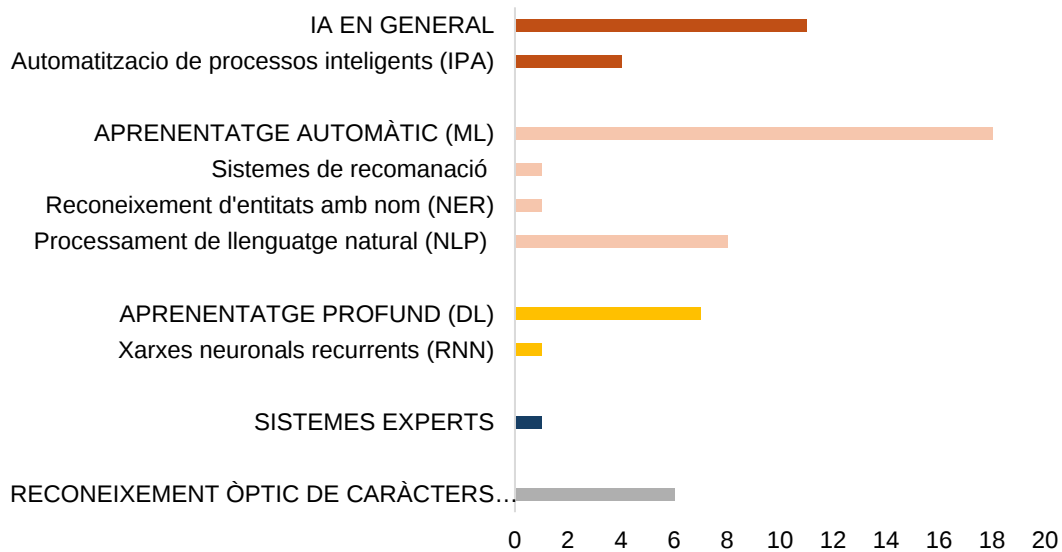


Fig. 8 Tipus d'IA analitzada o aplicada als estudis seleccionats (agrupades per tipus i especificant les tècniques utilitzades).

En segon lloc, s'ha analitzat el tipus de tractament de les dades segons l'àmbit arxivístic d'aplicació en cadascun dels estudis seleccionats. Aquests àmbits s'han identificat mitjançant cinc àmplies categories corresponents a diferents processos de gestió documental: creació i ús ($n=2$), gestió i administració ($n=13$), classificació i indexació ($n=2$), preservació ($n=11$), i accés ($n=8$). És important mencionar que, igual que en l'apartat anterior, molts estudis no es podien classificar dins una sola categoria, ja que incloïen diversos temes o àmbits d'aplicació.

En últim lloc, s'ha descrit el tipus d'organització pública en la qual els estudis seleccionats analitzen o apliquen l'ús de la IA. Degut a la gran varietat de denominacions per referir-se al mateix organisme, s'han classificat les diferents organitzacions públiques segons el seu àmbit geogràfic. L'anàlisi mostra que la majoria d'estudis prenen com abast les organitzacions nacionals (governamentals) ($n=22$) (p.e. (Allothman and Sait 2022, Azzam et al. 2023, Baron and Payne 2017), mentre que un nombre molt menor fa referència de manera genèrica al sector públic o a les organitzacions públiques ($n=3$) (Alaoui 2023, Duranti 2022, Jaillant and Rees 2023). A més, s'identifiquen altres aplicacions en l'àmbit regional ($n=2$) (Chabin 2020, Modiba, Ngoepe, and Ngulube 2019) i en l'àmbit local ($n=2$) (Hashimoto 2023, Chabin 2020).

4.5 Limitacions

La revisió d'abast (*scoping review*) realitzada en aquest treball ha seguit mètodes rigorosos i transparents al llarg de tot el procés. No obstant això, existeixen algunes limitacions que és necessari assenyalar. En primer lloc, cal tenir en compte que les

revisions d'abast sempre tenen certes limitacions inherents, ja que el seu objectiu és proporcionar informació àmplia sobre un tema, sense aprofundir en aspectes concrets. D'altra banda, és possible que en la revisió duta a terme no s'hagi identificat tota la literatura publicada, malgrat els esforços per ser el més exhaustiu possible. Tot i que la revisió inclou estudis en qualsevol idioma, els termes de cerca s'han utilitzat principalment en anglès i castellà. Per tant, és possible que no s'hagin detectat estudis d'interès en altres idiomes, més enllà dels localitzats en la revisió de les citacions bibliogràfiques, tot i ser conscients del prestigi de l'arxivística en dominis lingüístics com el francès o l'alemany. Tanmateix, en el marc d'un treball acadèmic d'aquest tipus s'ha considerat que la bibliografia cercada en anglès i castellà era prou representativa per assolir els objectius proposats.

El vocabulari utilitzat en les cerques es va identificar a partir de fonts confiables que tractaven el tema de la IA vinculada als arxius. Tenint en compte que un dels principals termes utilitzats en la cerca és "IA", "*artificial intelligence*" o "*inteligencia artificial*", tots els estudis que utilitzaven altres termes més específics poden haver quedat fora dels resultats. No obstant això, és probable que aquells on s'utilitza una terminologia tan específica, corresponguin a àmbits d'estudi fora de l'abast d'aquest treball, com les ciències de la computació o l'enginyeria.

Un altre aspecte relacionat és l'ús del terme "IA" a la literatura d'una manera poc transparent. Durant la cerca es va detectar un nombre d'estudis en què s'utilitzava el terme "IA" i els seus derivats en el títol, paraules clau i resums, però que posteriorment en fer una lectura completa del text, no tractaven específicament sobre IA. Per aquest motiu, un cop revisats íntegrament els estudis prèviament seleccionats, el nombre va disminuir considerablement.

En conseqüència, s'ha pogut evidenciar el que es va intuir inicialment. Existeix una bona quantitat d'estudis que tracten el tema de la IA a les administracions públiques, així com la IA en relació amb els arxius en general, i en particular amb els arxius patrimonials i històrics. Tanmateix, són pocs els estudis que incloguin les tres variables d'interès per aquest treball: arxius / IA / administracions públiques.

Per tot això, i considerant el limitat nombre de resultats obtinguts finalment, es considera que aquesta revisió ha de ser interpretada amb prudència, tenint en compte que es tracta d'un àmbit en constant evolució.

5. DISCUSSIÓ

Fa aproximadament una dècada, l'ús de la IA als arxius encara era majoritàriament experimental. No obstant això, ja mostrava un enorme capacitat per modificar certs aspectes de la tasca diària dels/les arxivers/es, especialment en àmbits específics com els arxius històrics o patrimonial (Colavizza et al. 2022). No seria fins uns anys més tard que l'interès per part de les administracions públiques en aquest nou model tecnològic augmentaria exponencialment, fins al punt de desenvolupar estratègies nacionals (Govern d'Espanya 2020), i regulacions per promoure i gestionar correctament l'ús de la IA (Unió Europea 2021).

Actualment, tal com es reflecteix en aquest treball, la relació entre els arxius i les noves tecnologies és una realitat en constant creixement, però que encara planteja dubtes i preguntes sense resposta entre els/les professionals de l'arxivística. Aspectes com l'ètica, la seguretat i privacitat de les dades, així com l'impacte social i econòmic, són només algunes consideracions que cal tenir en compte en la integració de la IA.

Aquest treball presenta una revisió sistematitzada (*scoping review*) sobre l'aplicació i ús de la IA als arxius de les administracions públiques, en la que s'han examinat 27 estudis rellevants. Els resultats obtinguts evidencien múltiples aspectes a tenir en compte. A continuació es presenten les evidències més significatives detectades en l'anàlisi relacionada amb el procés i el contingut. Aquestes pretenen servir com a guia per avançar en el desenvolupament de la IA als arxius, i especialment als arxius públics.

5.1. Elements relacionats amb el procés

1. Augment dels estudis en els últims anys

Els resultats obtinguts indiquen un augment notable en la producció científica sobre l'ús de la IA aplicada als arxius de les administracions públiques, especialment a partir de 2021. Aquest notable increment podria ser atribuït a diversos factors, com ara el continu progrés tecnològic, les cada vegada més àmplies aplicacions pràctiques, l'augment de *start-ups* i empreses d'IA, o el desenvolupament de polítiques i estratègies nacionals per fomentar el seu ús (Gross 2020). No obstant això, probablement dos dels factors clau són l'augment progressiu de la quantitat de dades generades i gestionades per les administracions públiques; i la sobtada aparició de la Covid-19 l'any 2020. Aquesta última va evidenciar la necessitat d'accelerar els processos digitals i trobar solucions tecnològiques innovadores, per gestionar i analitzar de manera efectiva el gran volum de dades de l'administració (Vinuesa et al. 2020).

2. Ampliar el nombre d'estudis d'investigació aplicada

La majoria dels estudis analitzats han estat realitzats mitjançant revisions bibliogràfiques i estudis de cas. Aquests dos mètodes de recerca són lleugerament predominants en comparació amb la resta, tenint en compte que els estudis d'investigació aplicada es troben a continuació. Una evidència molt positiva per al sector, ja que permet no tan sols adquirir un coneixement des d'un punt de vista teòric, sinó també aplicar els nous models tecnològics en casos pràctics. Tanmateix, és important continuar incrementant els casos d'aplicació pràctica en contextos reals, per tal d'evidenciar els beneficis i els riscos d'una tecnologia que avança ràpidament (Azzam et al. 2023, Hu et al. 2005, Gelashvili and Pappel 2021).

3. Diversificar l'enfocament metodològic de les investigacions

En paral·lel, els resultats obtinguts es presenten majoritàriament mitjançant un enfocament metodològic qualitatiu. Tan sols un nombre mínim d'estudis utilitzen una metodologia mixta, i en cap cas s'adopta un enfocament exclusivament quantitatiu. És important tenir en compte que per a resoldre els problemes metodològics relacionats amb la IA aplicada a qualsevol àmbit, cal adquirir objectivitat, rigor i precisió sobre els fenòmens i efectes que comporta. Tal com argumenta (Lindgren 2022), són necessaris més estudis quantitatius per ampliar el coneixement de la IA, i comprendre com aquesta tecnologia afecta la societat, més enllà dels suggeriments derivats d'investigacions qualitatives.

4. Fomentar noves pràctiques als països en desenvolupament

Els resultats obtinguts mostren que els territoris amb una major producció científica responen als anomenats països avançats, destacant especialment el Regne Unit i els EUA. En ambdós casos es tracta de potències econòmiques que actualment es troben al capdavant de la investigació en IA aplicada als arxius governamentals, superant en qualitat a potències tecnològiques com la Xina (Baker 2017). Un lideratge que no seria possible sense institucions com els Arxius Nacionals del Regne Unit (*The National Archives, TNA*), capdavantera en termes d'innovació tecnològica i ús de la IA en tasques com el processament de textos i OCR, l'anàlisi de dades i la classificació automàtica (The National Archives 2021, Jaillant, Aske, and Caputo 2024). De manera similar, l'Administració Nacional d'Arxius i Registres (NARA - EUA), contribueix significativament a la producció científica relacionada amb l'ús de la IA com a part de les noves estratègies de govern digital (Ingram and Johnson 2022).

En paral·lel observem altres casos, com el de Sud-àfrica, que mostra una activitat científica notable en aquest àmbit. Tanmateix, juntament amb altres països del continent africà com Nigèria, requereixen una major implicació dels seus governs en l'ús i desenvolupament de la IA (Modiba, Ngoepe, and Ngulube 2019, Omigie, Krubu, and Anthony 2023). Aspectes com la millora de les infraestructures tecnològiques, una major inversió en recerca i desenvolupament (I+D), així com la creació de polítiques de suport i legislació adequades (Modiba 2022b), són importants per fomentar una major producció científica, que al mateix temps es reflecteixi en aplicacions pràctiques que permetin millorar les tècniques de gestió dels arxius en països emergents.

5. Compartir àmpliament les dades d'investigació utilitzades en els estudis

Per impulsar una investigació més oberta i transparent sobre l'aplicació i ús de la IA, és necessari posar a disposició d'altres investigadors/es les dades no processades de la recerca. Una pràctica que s'ha detectat en molt pocs casos en la selecció d'estudis realitzada. L'anomenat "*data sharing*" permet maximitzar la utilitat de les dades, ja que poden ser utilitzades per noves investigacions, així com millorar i comparar models o hipòtesis prèviament creades, fomentant d'aquesta manera l'augment del coneixement (Sixto et al. 2019).

Tot i que és una pràctica que es va iniciar en l'àmbit de les investigacions clíniques (Eastman 2014), actualment s'està promovent a tota la comunitat científica, incloent-hi les humanitats (Arthur and Hearn 2024). Per tal que les dades puguin ser compartides, han de complir un conjunt de qualitats precises i mesurables conegudes com a principis FAIR (*Findable, Accessible, Interoperable, Reusable*) (Wilkinson et al. 2016). Així doncs, és fonamental potenciar i remarcar la importància d'aquesta tasca en qualsevol tipus d'organització o àmbit acadèmic, fomentant una investigació en què l'accés obert, la qualitat i la transparència siguin claus per augmentar el coneixement sobre la IA als arxius de les administracions públiques.

6. Ampliar la investigació conceptual i pràctica sobre IA als arxius de les AAPP

Les investigacions sobre l'aplicació i ús de la IA als arxius del sector públic encara són escasses. La IA és una tecnologia que s'està desenvolupant ràpidament en el sector privat, fomentant la creació de models adaptats majoritàriament a les necessitats d'aquest sector (Govern d'Espanya 2024). Tanmateix, aquests nous productes comercials són posteriorment adoptats per la majoria d'institucions arxivístiques públiques per millorar la seva eficiència en les tasques diàries (The National Archives 2021, Venkata et al. 2022, Ingram and Johnson 2022).

La manca d'investigació pràctica i conceptual aplicada directament al sector públic, provoca una sèrie de deficiències en aquests productes respecte a les necessitats específiques d'aquest sector. Això es tradueix en una manca de control i sobirania sobre les dades públiques, un elevat cost per adaptar els models correctament amb l'objectiu d'evitar decisions esbiaixades, i una manca de confiança relacionada amb la sensibilitat i privacitat de les dades (Venkata et al. 2022).

Per aquest motiu, és important fomentar estratègies per ampliar la investigació i desenvolupament intern de la IA adaptada als arxius de les administracions públiques. Una tasca que permetrà millorar l'eficiència i assegurar un major control sobre les dades públiques.

5.2 Elements relacionats amb el contingut

1. Àmplia consciència dels riscos ètics de la IA

En el conjunt d'estudis seleccionats s'ha detectat una gran àrea d'estudi relacionada amb els riscos de la IA (biaixos, discriminació, sexisme...) en l'àmbit dels arxius públics. Aquest fet s'ha d'interpretar com la presència d'un important marc de conscienciació respecte a les noves tecnologies i els arxius. Un aspecte molt important a l'hora de construir des de les primeres fases, models o noves eines tecnològiques que parteixin de la sostenibilitat, l'ètica i el respecte. Es tracta de qüestions que han pres molta força amb l'aparició de la IA en múltiples àmbits del nostre entorn (Dubber, Pasquale, and Das 2020), un fet que no podia passar per alt el sector arxivístic, i més concretament l'encarregat de la gestió i preservació de la informació pública. Tanmateix, tot i observar un augment en la preocupació i consciència dels riscos ètics de la IA, és rellevant posar èmfasi en la gestió pràctica d'aquests riscos, més enllà de la teoria.

Per aquest motiu, assegurar la creació de conjunts de dades sobre conceptes ètics (no discriminació, inclusió, consentiment...) (Jaillant, Aske, and Caputo 2024), supervisar adequadament els processos i la lògica subjacent en l'entrenament dels models (Alaoui 2023), i desenvolupar programes de capacitat ètica i comprensió conscient dels riscos entre els/les arxivers/es (Ingram and Johnson 2022), són alguns dels punts a tenir en compte per establir una pràctica segura, fiable i responsable (Generalitat de Catalunya 2020a).

2. Millora de les tasques diàries dels/les arxivers/es

La IA és capaç de transformar significativament les tasques diàries dels/les arxivers/es a les administracions públiques. Així ho evidencien els resultats

obtinguts, on els autors/es exposen els múltiples avantatges que comporta l'aplicació de noves formes de treball basades en IA en la majoria dels àmbits arxivístics (4.4.1). Això implica que els arxius, i especialment els/les arxivers/es, poden beneficiar-se de les capacitats de la IA en la major part de les seves tasques diàries, que inclouen des de la gestió de les dades fins a la preservació i accés a la informació. Tanmateix, els resultats obtinguts mostren un major interès en tasques relacionades amb la gestió i administració de les dades. En aquest àmbit, els/les autors/es destaquen les possibilitats de la IA per crear i modernitzar sistemes de gestió documental (Stylianou et al. 2022), que permetin millorar la gestió interna de manera més precisa (Gelashvili and Pappel 2021), reduint els errors i millorant la coherència en la gestió (Omigie, Krubu, and Anthony 2023). El creixent volum de dades creades i emmagatzemades entorn de les administracions públiques, promou la necessitat d'una gestió eficient i segura de la informació (Lindgren 2022). Per aquest motiu, la integració de la IA en tasques diàries com la classificació, l'avaluació i l'organització dels documents millora la gestió de les dades i la informació, augmentant l'eficiència, la precisió i la seguretat en el maneig dels arxius públics.

3. Millorar en la concreció sobre el tipus d'IA utilitzada

En una part dels estudis analitzats l'enfocament sobre el tipus d'IA utilitzada és força concret, on es mencionen tècniques específiques com el processament de llenguatge natural (NLP) o les xarxes neuronals recurrents (RNN). No obstant això, es detecta que la major part dels estudis utilitzen una denominació molt àmplia. En conseqüència, trobem una gran quantitat de referències a termes com aprenentatge automàtic, que si bé coneixem el seu funcionament, en pocs casos s'aprofundeix més en les seves tipologies (supervisat, no supervisat, semisupervisat). En la revisió duta a terme, només dos casos (Hu et al. 2005, Reutter and Spiiker 2019) especificaven clarament quin enfocament s'utilitzava dins el mateix model. De la mateixa manera, s'observa que el terme "IA" s'utilitza de forma genèrica en diversos estudis, una pràctica que és important evitar, ja que dificulta les comparacions i les metaanàlisis entre estudis en diferents contextos, àrees i períodes. Malgrat això, observem que l'aprenentatge automàtic continua sent la forma d'aplicació predominant en relació amb la resta de subconjunts propis de la IA. Un fet impulsat per la seva capacitat de gestionar grans volums de dades en processos mecànics, però que al mateix temps requereix un mínim control humà, cosa que el fa especialment adequat per a les característiques pròpies del sector públic.

4. Incrementar els estudis a altres tipus d'organitzacions públiques

La majoria d'estudis se centren en tipologies d'organitzacions molt concretes, de manera que la inespecificitat en les àrees o dominis s'observa en casos molt residuals i, generalment, atribuïble al caràcter teòric de l'estudi. No obstant això, el tipus d'organització pública d'aplicació majoritària, són les que fan referència a l'àmbit nacional (agències i departaments governamentals, arxius nacionals i altres entitats geogràficament similars). Són pocs els estudis que analitzen l'ús de la IA en contextos regionals o locals (Hashimoto 2023, Chabin 2020, Modiba, Ngoepe, and Ngulube 2019). Per això, l'increment d'estudis aplicats a menor escala pot ser beneficiós, donades les necessitats i desafiaments específics que difereixen dels nivells nacionals. Estudiar l'ús i aplicació de la IA en aquests contextos permet desenvolupar solucions adaptades i més eficients a llarg termini, que poden ser extrapolables a altres organitzacions amb característiques similars.

5. Més implicació i proactivitat per part de la professió arxivística

És indubtable afirmar que l'aparició de la IA en el camp de l'arxivística modificarà el rol professional dels/les arxivers/es (Theimer 2018), i en conseqüència, el paper dels arxius com a institució (Reutter and Spíker 2019). El seu paper tradicional, basat en la gestió i custòdia de la informació, es modificarà activament per a convertir-se en professionals experts/es en dades (Cushing and Osti 2023). De moment, és important verificar si l'actitud positiva del sector públic cap a la IA, tal com demostren alguns estudis, és realment tan comuna (Lindgren 2022). Tot i això, és necessari que la professió ofereixi serveis i experiències conjunes amb altres professions afins. Una tasca que no és possible dur a terme sense la col·laboració interdisciplinària. Són múltiples els/les autors/es que subratllen la necessitat de crear estratègies i processos de participació conjunta i assessorament amb altres sectors especialitzats, com els/les enginyers/es de dades, programadors/es, i altres professionals tecnològics, per promoure el desenvolupament de solucions adaptades a les necessitats dels arxius (Jaillant and Rees 2023, Jaillant, Aske, and Caputo 2024, Colavizza et al. 2022).

De la mateixa manera, diversos estudis exploren el potencial dels productes comercials basats en eines d'IA per la gestió dels documents d'arxiu (The National Archives 2021, Venkata et al. 2022, Hutchinson 2020). Aquests conclouen en què és necessari continuar experimentant de forma pràctica amb aquests productes, i fomentar al mateix temps, una col·laboració activa entre arxivers/es i empreses tecnològiques proveïdores d'eines d'IA (Jaillant 2022b). Una tasca que no només

afavoreix l'adaptació de les eines a les necessitats del sector, sinó que pot millorar aspectes tan rellevants per al sector públic com són la privacitat, l'ètica i la seguretat de la informació.

6. Investigar i desenvolupar pràctiques de gestió de riscos als arxius de les AAPP

En els estudis seleccionats s'observa una clara preocupació envers els riscos associats a l'ús de la IA ([4.4.1](#)). Alguns dels riscos més habituals són els biaixos i la discriminació induïts pels models algorítmics, així com la manca de transparència i explicabilitat (XAI). No obstant això, són pocs els/les autors/es que proposen pràctiques concretes per abordar aquests riscos, dels quals els/les arxivistes/es en són plenament conscients.

Alguns projectes com (The National Archives 2021) exploren les implicacions ètiques de l'ús de la IA i proposen marcs per una implementació responsable. Altres, com (Venkata et al. 2022) fan referència a les diverses iniciatives governamentals centrades en la creació de marcs reguladors, abordant especialment la gestió dels riscos associats. El coneixement sobre com poden impactar les noves solucions basades en IA a les organitzacions públiques, i més concretament als arxius, és sense dubte una tasca ja iniciada amb la identificació i conscienciació dels riscos. Tanmateix, és necessari continuar investigant en contextos localitzats per comprendre millors com afrontar-los (Jaillant and Caputo 2022). Un increment en la recerca sobre pràctiques de gestió de riscos adequadament contextualitzades i basades en aspectes com: el desenvolupament de polítiques i protocols de seguretat, formació en gestió de riscos, i monitoratge i actualització continua per detectar i fer front a noves amenaces, són alguns punts a tenir en compte per gestionar adequadament els riscos i promoure un ús ètic i segur de la IA als arxius.

7. Investigar com mesurar l'impacte de les solucions d'IA als arxius de les AAPP

En la selecció d'estudis, diversos autors/es mostren una preocupació sobre l'impacte que tindrà la IA als arxius públics (p.e. (Colavizza et al. 2022, Stylianou et al. 2022, Azzam et al. 2023)). Alguns d'ells reconeixen que l'impacte social i econòmic de les noves solucions d'IA encara és incert i, per tant, és necessari continuar avaluant-lo (Azzam et al. 2023).

Un primer pas és revisar com aquesta preocupació es planteja en la bibliografia acadèmica, i així adoptar una millor comprensió del seu impacte (Colavizza et al. 2022). L'avaluació mitjançant l'anàlisi d'estudis de cas, juntament amb entrevistes i enquestes (Gelashvili and Pappel 2021) que permetin recollir opinions i percepcions dels/les professionals del sector, així com experimentar de manera pràctica amb

productes comercials (Venkata et al. 2022), són enfocaments clau per conèixer millor l'impacte positiu i negatiu de la IA. Tot i que aquest impacte encara s'està mesurant en la major part dels sectors econòmics i socials (Govern d'Espanya 2024), cal tenir en compte que les afectacions en el cas del sector públic poden ser molt àmplies. Per aquest motiu, és fonamental continuar investigant com mesurar l'impacte de la IA en la presa de decisions i la transparència de les organitzacions públiques (Lindgren 2022), i més concretament als arxius, degut a la important responsabilitat social que tenen sobre la informació pública.

8. Augmentar la inversió i el compromís de les administracions públiques

Les administracions públiques tenen un paper fonamental en l'aplicació i ús de la IA als arxius públics. Per aconseguir una implementació real i efectiva de les noves eines tecnològiques, és necessari que les administracions en fomentin la presència, no només en les tasques administratives i d'atenció al ciutadà, sinó també en els seus procediments interns.

En aquest sentit, és important destinar les inversions adequades amb l'objectiu de dotar de les infraestructures tecnològiques necessàries per produir resultats òptims, tal com remarquen alguns dels estudis seleccionats (Reutter and Spíker 2019). De la mateixa manera, destinar recursos a la formació i desenvolupament de noves competències, enfocades a capacitar els/les professionals dels arxius en l'ús i desenvolupament de noves eines basades en IA, és una de les demandes més freqüents dels/les autors/es (Rolan et al. 2019, Ingram and Johnson 2022, The National Archives 2021). Una formació adequada permetrà als equips de gestió documental adquirir coneixements sobre en els aspectes tècnics i ètics de la IA, millorant no només els coneixements, sinó la confiança envers les noves eines. Això es traduirà en un major ús i, per tant, en una optimització de les tasques habituals (Gelashvili and Pappel 2021). Per tot això, és fonamental que les administracions públiques promoguin i avaluin projectes de recerca i projectes pilot, que explorin l'ús concret de la IA en les diverses tasques arxivístiques. En conjunt, augmentar la inversió i promoure la col·laboració amb altres sectors (Venkata et al. 2022), proporcionarà informació determinant per impulsar un major desenvolupament de les infraestructures digitals en l'àmbit dels arxius públics.

6. CONCLUSIONS

En aquest treball es presenta una anàlisi contextual sobre el coneixement i desenvolupament de la IA aplicada a la gestió dels arxius en l'àmbit de les administracions públiques. Aquesta anàlisi s'ha dut a terme mitjançant una revisió sistematitzada, concretament una revisió d'abast o *scoping review*, de la literatura existent. Els resultats obtinguts mitjançant l'anàlisi de 27 estudis seleccionats, juntament amb una revisió del marc teòric, ha permès identificar aspectes clau a tenir en compte sobre el context actual. Finalment, s'han evidenciat sis aspectes rellevants relacionats amb el procés, i vuit relacionats amb el contingut dels estudis. Aquests elements permeten comprendre l'estat actual de l'aplicació i ús de la IA als arxius del sector públic, així com guiar el desenvolupament de solucions eficients per abordar les principals problemàtiques identificades.

El conjunt d'evidències obtingudes mostra una dissonància entre la investigació i les necessitats pràctiques. Per un costat, la investigació conceptual i metodològica sobre la IA aplicada als arxius del sector públic ha augmentat notablement en els darrers anys, oferint un marc d'anàlisi sobre com utilitzar i avançar en l'ús d'aquestes noves eines per al desenvolupament i millora de les pràctiques arxivístiques. Malgrat això, són pocs els estudis empírics duts a terme per part del sector arxivístic, essent la majoria realitzats per altres disciplines relacionades. Per això, es considera necessari que la investigació futura fomenti estratègies per ampliar la recerca i desenvolupament intern de la IA adaptada als arxius de les administracions públiques. D'aquesta manera, el suport empíric i les implementacions pràctiques permetran considerar aspectes únics dels arxius públics, mitjançant l'aplicació i difusió d'un marc específic.

Un altre idea clau, relacionada amb l'anterior, és la consciència sobre aspectes com la confiança, la transparència i el desenvolupament de solucions explicables, així com les preocupacions ètiques al voltant de la privacitat i protecció de les dades. Es tracta d'un tema àmpliament abordat en la literatura analitzada i, per tant, s'evidencia com una de les principals inquietuds dels/les professionals dels arxius. A més, en l'entorn del sector públic, els aspectes ètics adquireixen una dimensió addicional, ja que la privacitat i protecció de les dades es troba sotmesa a un compliment legal i a una responsabilitat pública amb la ciutadania, garantint-ne un tractament responsable i segur. Per aquest motiu, la presència dels/les arxivers/es en les primeres fases de creació de les eines basades en IA és fonamental. Proporcionar les dades adequades als algoritmes permetrà garantir que les decisions automatitzades futures siguin més justes i

transparents, evitant els biaixos racials, de gènere, de classe..., i fomentant al mateix temps el desenvolupament de solucions explicables i interpretables.

Per tal d'assolir la situació exposada, és fonamental la responsabilitat i participació activa de les institucions públiques. La promoció de polítiques al voltant de les noves eines d'IA, aplicades no només a aspectes relacionats directament amb els/les usuaris/es, sinó també als procediments interns de treball, permetrà facilitar la transparència i accés a la informació pública, garantint al mateix temps la privacitat de les dades. La creació i manteniment de portals de transparència, on els/les ciutadans/es poden accedir fàcilment a la informació pública i a les dades obertes, són eines importants per generar confiança en la ciutadania, i fomentar la seva participació en la presa de decisions. Per tant, la dotació de recursos humans, econòmics i tecnològics és essencial per produir resultats òptims que garanteixin la confiança, transparència, privacitat i seguretat, actuant sota una responsabilitat ètica i promovent la construcció d'una ciutadania més confiada i informada.

De la mateixa manera, els/les professionals dels arxius tenen un paper fonamental en la correcta execució d'aquest nou entorn tecnològic. Per aquest motiu, la capacitació de l'equip sobre nous conceptes i pràctiques relacionades amb IA, així com la participació en programes de formació contínua per mantenir-se actualitzats, són aspectes molt importants a considerar. Per assolir aquests objectius correctament, la col·laboració amb altres professionals interdisciplinaris, així com la participació en xarxes professionals basades en l'intercanvi de coneixements, permetran millorar les pràctiques i contribuir al desenvolupament d'estratègies afins. En conjunt, la responsabilitat actual dels arxiviers implica una combinació de competències tècniques, legals, ètiques i de compromís amb la transparència i seguretat, que indubtablement estan transformant el rol professional i tradicional del sector. Per tant, cal considerar i ser conscients de l'impacte que tenen les decisions preses en la gestió de les dades en un context tecnològic com l'actual, i en conseqüència, treballar en la implementació de pràctiques sostenibles, promovent una representació i un accés més just i equitatiu a la informació.

En definitiva, la recerca duta a terme pretén aportar més coneixement sobre la investigació i aplicació actual de les eines d'IA als arxius de les administracions públiques. En ell es destaca la importància creixent de la integració d'aquestes noves tecnologies en la gestió dels arxius, tot posant de manifest una sèrie d'ítems a tenir en compte per assolir una implementació responsable i conscient. Mitjançant l'anàlisi de la literatura existent, l'estudi pretén ser una eina útil com a base per a futures

investigacions i per a la millora continua de solucions eficients que garanteixin l'ús ètic i responsable de la IA en la gestió dels arxius, assegurant la integritat, la seguretat i la transparència de les dades.

Declaració de disponibilitat de dades

Les dades que donen suport als resultats obtinguts en aquest treball acadèmic, es troben disponibles a <https://doi.org/10.6084/m9.figshare.26324779>.

7. BIBLIOGRAFIA

- Aangenendt, Gijs. 2022. "Archives in the Digital Age : The use of AI and machine learning in the Swedish archival sector." Independent thesis Advanced level (degree of Master (Two Years)) Student thesis, Theses within Digital Humanities (17).
- Agarwal, P. K. 2018. "Public Administration Challenges in the World of AI and Bots." 78 (6):917-921. doi: 10.1111/puar.12979.
- Alaoui, S. 2023. "Artificial intelligence and records management: what gains? What stakes?" *Canadian Journal of Information and Library Science-Revue Canadienne Des Sciences De L Information Et De Bibliotheconomie* 46 (2):20. doi: 10.5206/cjils-rcsib.v46i2.16695.
- Allothman, Abdulaziz Fahad, and Abdul Rahaman Wahab Sait. 2022. "Managing and retrieving bilingual documents using artificial intelligence-based ontological framework." *Computational Intelligence and Neuroscience* 2022:15. doi: 10.1155/2022/4636931.
- Androutsopoulou, Aggeliki, Nikos Karacapilidis, Euripidis Loukis, and Yannis Charalabidis. 2019. "Transforming the communication between citizens and government through AI-guided chatbots." *Government Information Quarterly* 36 (2):358-367. doi: 10.1016/j.giq.2018.10.001.
- Antman, Elliott M., Joseph Lau, Bruce Kupelnick, Frederick Mosteller, and Thomas C. Chalmers. 1992. "A Comparison of Results of Meta-analyses of Randomized Control Trials and Recommendations of Clinical Experts: Treatments for Myocardial Infarction." *JAMA* 268 (2):240-248. doi: 10.1001/jama.1992.03490020088036.
- Archives, The National. 2021. Using AI for Digital Records Selection in Government. Guidance for records managers based on an evaluation of current marketplace solutions.
- Arksey, Hilary, and Lisa O'Malley. 2005. "Scoping studies: towards a methodological framework." *International Journal of Social Research Methodology* 8 (1):19-32. doi: 10.1080/1364557032000119616.
- ARMA, Association of Records Managers and Administrators. 2017. Principios de Mantenimiento de Registros Generalmente Aceptados.
- Arthur, Paul Longley, and Lydia Hearn. 2024. *Open Scholarship in the Humanities*: Bloomsbury Academic.
- Associates, Cohasset, and ARMA International. 2019. Information Governance Benchmarking Report.
- Aydın, Cengiz. 2023. "Dijital Dönüşüm Bağlamında Belgenin Yaşam Döngüsü Süreçlerinin Yönetiminde Yapay Zekâ Kullanımı." *Bilgi Yönetimi* 6 (1):21-42. doi: 10.33721/by.1268483.
- Azzam, F., M. Jaber, A. Saies, T. Kirresh, R. Awadallah, A. Karakra, H. Barghouthi, and S. Amarnah. 2023. "The Use of Blockchain Technology and OCR in E-

- Government for Document Management: Inbound Invoice Management as an Example." *Applied Sciences (Switzerland)* 13 (14). doi: 10.3390/app13148463.
- Baker, Simon. 2017. "Which countries and universities are leading on AI research?". *Times Higher Education*, accessed 12 juny. <https://www.timeshighereducation.com/data-bites/which-countries-and-universities-are-leading-ai-research>.
- Baron, J. R., and N. Payne. 2017. "Dark archives and edemocracy: Strategies for overcoming access barriers to the public record archives of the future." *Proceedings of the 7th International Conference for E-Democracy and Open Government, CeDEM 2017, Krems, Austria*.
- Bauman, Zygmunt. 2001. *Globalització: Les conseqüències humanes*: Edicions de la Universitat Oberta de Catalunya.
- Bazán-Gil, Virginia. 2023. "Artificial intelligence applications in media archives." *Profesional de la informació / Information Professional* 32 (5). doi: 10.3145/epi.2023.sep.17.
- Bell, Mark. 2020. "From tree to network: reordering an archival catalogue." *Records Management Journal* 30 (3):379-394. doi: 10.1108/RMJ-09-2019-0051.
- Benson, Allen C. 2009. "The Archival Photograph and Its Meaning: Formalisms for Modeling Images." *Journal of Archival Organization* 7 (4):148-187. doi: 10.1080/15332740903554770.
- Berning Prieto, Antonio David. 2023. "El uso de sistemas basados en inteligencia artificial por las Administraciones públicas: estado actual de la cuestión y algunas propuestas ad futurum para un uso responsable." *Revista de Estudios de la Administración Local y Autonómica* (20):165-185. doi: 10.24965/reala.11247.
- Binmakhashen, Galal M. , and Sabri A. Mahmoud. 2019. "Document Layout Analysis: A Comprehensive Survey." *ACM Comput. Surv.* 52. doi: 10.1145/3355610.
- Birkland, Thomas A. 2019. *An introduction to the policy process: Theories, concepts, and models of public policy making*: Routledge.
- Bocyt, Rasa, and Johan Oomen. 2020. "Content Adaptation, Personalisation and Fine-grained Retrieval: Applying AI to Support Engagement with and Reuse of Archival Content at Scale." *ICAART* (1).
- Booth, Andrew, Diana Papaioannou, and Anthea Sutton. 2012. *Systematic Approaches to a Successful Literature Review*.
- Bright, Jonathan, Bharath Ganesh, Cathrine Seidelin, and Thomas M Vogl. 2019. "Data science for local government."
- Bushey, J. 2023. "AI-Generated Images as an Emergent Record Format." 2023 IEEE International Conference on Big Data (BigData), 15-18 Dec. 2023.
- Cameron, Scott , Pat Franks, and Babak Hamidzadeh. 2023. "Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts." 16 (Comput. Cult. Herit.). doi: 10.1145/3594728.

- Campos Martínez, Maria del Pilar Campos. 2023. "Conjunts de dades, gestió documental i arxius: proposta de tractament." *Lligall* (46):131-145.
- Capellades, Alan. 2023. "Un matí al Port Aventura de les Dades. Els professionals de l'arxivística i la gestió de documents s'han reunit per abordar el tractament de les dades en un món governat pel Big Data i la Intel·ligència Artificial." *Núvol*, 5/5/2023.
- Casadesús de Mingo, Anahí. 2019. "La gestión del riesgo aplicada a la gestión de documentos y su impacto en la rendición de cuentas pública." Research thesis, Biblioteconomia. Documentació (Ciències Humanes), Universitat Autònoma de Barcelona.
- Casellas Serra, Lluís-Esteve. 2019. "La evaluación archivística en el nuevo contexto de las organizaciones." *La destrucció d'informació pública. Una mirada multidisciplinària sobre l'eliminació ordenada de la documentació*
- Catalunya, Generalitat de. 2014. Llei 19/2014, del 29 de desembre, de transparència, accés a la informació pública i bon govern. In *DOGC 6780*, edited by Generalitat de Catalunya. Barcelona.
- Catalunya, Generalitat de. 2020a. CATALONIA.AI L'Estratègia d'Intel·ligència Artificial de Catalunya. Barcelona: Departament de Polítiques Digitals i Administració Pública. Generalitat de Catalunya.
- Catalunya, Generalitat de. 2020b. DECRET 76/2020, de 4 d'agost, d'Administració digital. In *DOGC 8195*, edited by Generalitat de Catalunya. Barcelona.
- Catalunya, Generalitat de. 2023. ACORD GOV/158/2023, de 25 de juliol, pel qual s'aprova el Model de govern de les dades de l'Administració de la Generalitat de Catalunya i del seu sector públic. In *DOGC 8967*, edited by Generalitat de Catalunya. Barcelona.
- Cerrillo Martínez, Agustí. 2019. "Com obrir les caixes negres de les Administracions públiques? Transparència i rendició de comptes en l'ús dels algorismes (CA-EN)." *Revista catalana de dret públic*:13-28.
- Chabin, M. A. 2020. "The potential for collaboration between AI and archival science in processing data from the French great national debate." *Records Management Journal* 30 (2):241-252. doi: 10.1108/RMJ-08-2019-0042.
- Chronos, O., and S. Sundell. 2011. "Digitalkoot: Making old archives accessible using crowdsourcing." AAAI Workshop - Technical Report.
- Chun, A. H. W. 2007. "Using AI for e-Government automatic assessment of immigration application forms." Proceedings of the National Conference on Artificial Intelligence.
- Codina, Lluís. 2024. "Revisiones tradicionales, sistemáticas o de alcance: ¿cómo elegir el tipo de revisión de la literatura que corresponde en cada caso?" *Infonomy* 2 (2). doi: 10.3145/infonomy.24.021.
- Codina, Lluís, Carlos Lopezosa, and Pere Freixa Font. 2021. "Scoping reviews en trabajos académicos en comunicación: frameworks y fuentes." Larrondo Ureta A, Meso Ayerdi K, Peña Fernández S, editores. Información y Big Data en el

- sistema híbrido de medios-XIII Congreso Internacional de Ciberperiodismo País Vasco [Leioa]: Universidad del País Vasco, 15-17 nov 2021.
- Colavizza, G., T. Blanke, C. Jeurgens, and J. Noordegraaf. 2022. "Archives and AI: An Overview of Current Debates and Future Perspectives." *Journal on Computing and Cultural Heritage*.
- Colavizza, Giovanni, Maud Ehrmann, and Fabio Bortoluzzi. 2019. "Index-Driven Digitization and Indexation of Historical Archives." 6. doi: 10.3389/fdigh.2019.00004.
- Couture, Beatrice, Farah Verret, Maxime Gohier, and Dominique Deslandres. 2023. "The Challenges of HTR Model Training: Feedback from the Project Donner le gout de l'archive a l'ere numerique." *Journal of Data Mining Digital Humanities*.
- Criado-Pérez, Caroline. 2020. *La mujer invisible: descubre cómo los datos configuran un mundo hecho por y para los hombres*: Seix Barral.
- Criado, J. Ignacio. 2021. "Inteligencia Artificial (y Administración Pública)." *EUNOMÍA. Revista en Cultura de la Legalidad* (20):348-372. doi: 10.20318/eunomia.2021.6097.
- Cushing, Amber L, and Giulia Osti. 2023. ""So how do we balance all of these needs?": how the concept of AI technology impacts digital archival expertise." *Journal of Documentation* 79 (7):12-29.
- Dahl, Christian M., Torben S. D. Johansen, Emil N. Sørensen, Christian E. Westermann, and Simon Wittrock. 2023. "Applications of machine learning in tabular document digitisation." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 56 (1):34-48. doi: 10.1080/01615440.2023.2164879.
- Davet, J., B. Hamidzadeh, and P. Franks. 2023. "Archivist in the machine: paradata for AI-based automation in the archives." *Archival Science* 23 (2):275-295. doi: 10.1007/s10502-023-09408-8.
- Du, Lin , Brandon Le, and Edouardo Honig. 2024. "Probing Historical Image Contexts: Enhancing Visual Archive Retrieval through Computer Vision." 16 (Comput. Cult. Herit.). doi: 10.1145/3631129.
- Dubber, Markus D., Frank Pasquale, and Sunit (Eds.) Das. 2020. *The Oxford Handbook of Ethics of AI*: Oxford University Press.
- Duranti, Luciana. 2022. "Why a world gone digital needs archival theory more than ever before?" *Archeion* (123):1-21. doi: 10.4467/26581264ARC.22.015.16668.
- Duranti, Luciana, Muhammad Abdul-Mageed, Darra Hofman, and Peter Sullivan. 2021. "I Trust AI, el Nuevo Proyecto de investigación de InterPARES." *Anuario: Escuela de Archivología* (13):36-55.
- Eastman, Peggy. 2014. IOM Attempting to Set Principles for Responsible Clinical Trial Data Sharing. LWW.
- Espanya. 2013. Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno. In *BOE* 295, edited by Gobierno de España. Madrid.

- Espanya, Ministerio de Asuntos Económicos y Transformación Digital. 2020. Estrategia nacional de inteligencia artificial.
- Espanya, Ministerio de Asuntos Económicos y Transformación Digital. Gobierno de España. 2023. España digital: Informe sobre los avances en la Estrategia Nacional de Inteligencia Artificial 2026. In *Estrategia Nacional de Inteligencia Artificial (ENIA)*.
- Espanya, Ministerio para la Transformación Digital y de la Función Pública. 2024. Estrategia de Inteligencia Artificial 2024. edited by Ministerio para la Transformación Digital y de la Función Pública.
- Europea, Unió. 2021. Reglamento (UE) 2021/694 del Parlamento Europeo y del Consejo de 29 de abril de 2021 por el que se establece el Programa Europa Digital y por el que se deroga la Decisión (UE) 2015/2240 (Texto pertinente a efectos del EEE). In 166, edited by Consejo de la Unión Europea Parlamento Europeo.
- Europea, Unió. 2022. Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos). In 152, edited by Consejo de la Unión Europea Parlamento Europeo.
- Firmani, D., P. Merialdo, E. Nieddu, and S. Scardapane. 2017. "In codice ratio: OCR of handwritten Latin documents using deep convolutional networks." CEUR Workshop Proceedings.
- Fornes, Alicia, Veronica Romero, Arnau Baro, Juan Ignacio Toledo, Joan Andreu Sanchez, Enrique Vidal, and Josep Lladós. 2017. "ICDAR2017 Competition on Information Extraction in Historical Handwritten Records." 14th International Conference on Document Analysis and Recognition.
- Frankish, Keith, and William M. (Eds.) Ramsey. 2014. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press.
- Gasser, U., and V. A. F. Almeida. 2017. "A Layered Model for AI Governance." *IEEE Internet Computing* 21 (6):58-62. doi: 10.1109/MIC.2017.4180835.
- Gelashvili, T., and I. Pappel. 2021. "Challenges of transition to paperless management: Readiness of incorporating AI in decision-making processes." 2021 8th International Conference on eDemocracy and eGovernment, ICEDEG 2021.
- Gilliland, Anne J 2016. "Designing Expert Systems for Archival Evaluation and Processing of Computer-Mediated Communications." In *Archival Multiverse*, edited by A.J. Gilliland, S. McKemmish and A.J Lau, eds., 685-721. Melbourne: Monash University Press.
- Goudarouli, Eirini, Anna Sexton, and John Sheridan. 2019. "The Challenge of the Digital and the Future Archive: Through the Lens of The National Archives UK." *Philosophy & Technology* 32 (1):173-183. doi: 10.1007/s13347-018-0333-3.
- Gross, Kilian. 2020. WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. Poland: European Border and Coast Guard Agency

- Hashimoto, Yo. 2023. *Use of AI Tools in Local Governments and its Implications on Public Records Management*. Vancouver, BC: 4th International Symposium, ITrustAI.
- Hirtle, P.B. 1987. "Artificial Intelligence, Expert Systems, and Archival Automation." *Provenance, Journal of the Society of Georgia Archivists* 5 (1):8.
- Hu, Xiao, Larry S Jackson, Sai Deng, and Jing Zhang. 2005. "Automatic subject heading assignment for online government publications using a semi-supervised machine learning approach." *American Society for Information Science and Technology*, 18 October 2006.
- Hutchinson, Tim. 2020. "Natural language processing and machine learning as practical toolsets for archival processing." *Records Management Journal* 30 (2):155-174. doi: 10.1108/RMJ-09-2019-0055.
- Iberoamericano, Centro Cochrane. 2012. *Manual Cochrane de Revisiones Sistemáticas de Intervenciones (Versión 5.1.0)*. Cochrane London, UK.
- Ingram, William A, and Sylvester A Johnson. 2022. *Ensuring Scholarly Access to Government Archives and Records*. edited by Virginia Tech: Virginia Tech University Libraries, Virginia Tech Center for Humanities, U.S. National Archives and Records Administration.
- ISO, International Organization for Standardization. 2016. ISO 15489-1:2016 "Information and documentation. Records management" Part 1: Concepts and principles.
- ISO, International Organization for Standardization. 2022. ISO 24143:2022 "Information and documentation — Information Governance — Concept and Principles".
- Jaillant, L., K. Aske, and A. Caputo. 2024. Case study: The National Archives (UK). In *Navigating AI for Cultural Heritage Organisations*, edited by Jaillant L.; Warwick C.; Gooding P.; Aske K.; Layne-Worthey G.; and Downie JS: Loughborough University.
- Jaillant, L., and A. Rees. 2023. "Applying AI to digital archives: trust, collaboration and shared professional ethics." *Digital Scholarship in the Humanities* 38 (2):571-585. doi: 10.1093/llc/fqac073.
- Jaillant, Lise. 2022a. "How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions." *Archival Science* 22 (3):417-436. doi: 10.1007/s10502-022-09390-7.
- Jaillant, Lise. 2022b. "Introduction." In *Archives, Access and Artificial Intelligence*, edited by Lise Jaillant, 7-28. Bielefeld: transcript Verlag.
- Jaillant, Lise, and Katherine Aske. 2024. "Are Users of Digital Archives Ready for the AI Era? Obstacles to the Application of Computational Research Methods and New Opportunities." *Journal on Computing and Cultural Heritage* 16:Article 87. doi: 10.1145/3631125.
- Jaillant, Lise, and Annalina Caputo. 2022. "Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data." *AI & SOCIETY* 37 (3):823-835. doi: 10.1007/s00146-021-01367-x.

- Janssen, Marijn, and Natalie Helbig. 2018. "Innovating and changing the policy-cycle: Policy-makers be prepared!" *Government Information Quarterly* 35:S99-S105. doi: doi.org/10.1016/j.giq.2015.11.009.
- Jarvinen, Zachary. 2020. *Enterprise AI for dummies*. First ed. Hoboken (NJ). EUA: John Wiley & Sons Inc.
- Jo, Eun Seo , and Timnit Gebru. 2020. "Lessons from archives: strategies for collecting sociocultural data in machine learning." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain.
- Kampffmeyer, Ulrich. 2023. "Governança de la informació: un manual." *Lligall* (46):15-47.
- Kugley, Shannon, Anne Wade, James Thomas, Quenby Mahood, Anne-Marie Klint Jørgensen, Karianne Hammerstrøm, and Nila Sathe. 2016. "Searching for studies: A guide to information retrieval for Campbell." *Campbell Systematic Reviews* 10.
- Lee, C. A. 2018. "Computer-Assisted Appraisal and Selection of Archival Materials." 2018 IEEE International Conference on Big Data (Big Data), 10-13 Dec. 2018.
- Lindgren, Joakim. 2022. "Det våras för Eu Och Ai: Europeiska Kommissionens Rättsakt för Artificiell Intelligens / Springtime for EU and AI: The European commission's Artificial intelligence act." Master's thesis, Institutionen för ABM Uppsatser inom arkivvetenskap Uppsala Universitet (Sweden) (207).
- Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. *Big data: The next frontier for innovation, competition, and productivity*. Washington: McKinsey Global Institute.
- Mas, Joan, Alicia Fornes, and Josep Lladós. 2016. "An Interactive Transcription System of Census Records using Word-Spotting based Information Transfer." 12th IAPR Workshop on Document Analysis Systems.
- Mat Isa, Azman, and Nordiana Mohd Nordin. 2013. "Strategic records and risk management for the sustainability of organizations: a case study investigation." *Comma* 2013:29-40. doi: 10.3828/comma.2013.1.3.
- McCarthy, John 1956. "The Inversion of Functions Defined by Turing Machines." In *Automata Studies*. (AM-34), Volume 34, edited by C. E. Shannon and J. McCarthy, 177-182. Princeton: Princeton University Press.
- Milleville, Kenzo , Alec Van den Broeck, Nastasia Vanderperren, Rony Vissers, Matthias Priem, Nico Van de Weghe, and Steven Verstockt. 2023. "Enriching Image Archives via Facial Recognition." *Journal on Computing and Cultural Heritage* 16. doi: 10.1145/3606704.
- Misuraca, Gianluca, and Colin Van Noordt. 2020. "AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU." *JRC Research Reports* (JRC120399).
- Modiba, Mashilo. 2022a. "Artificial intelligence for the improvement of records management activities at the Council for Scientific and Industrial Research." *Journal of the South African Society of Archivists* 55:16-26.

- Modiba, Mashilo. 2022b. "Legislation Used to Apply Artificial Intelligence for the Management of Records at the Council for Scientific and Industrial Research in South Africa." *African Journal of Library Archives and Information Science* 32 (1):21-35.
- Modiba, Thomas, Mpho Ngoepe, and Patrick Ngulube. 2019. "Application of disruptive technologies to the management and preservation of records." *Mousaion: South African Journal of Information Studies* 37 (1):14. doi: doi.org/10.25159/2663-659X/6159.
- Mordell, Devon. 2023. *Traçant un nou paradigma arxivístic: (Arxius com a) Dades*. 19è Congrés d'Arxivística i Gestió de Documents de Catalunya. Sant Cugat del Vallès.
- Moss, Michael S, and Tim J Gollins. 2017. "Our digital legacy: an archival perspective." In *Michael Moss on Archives*, 160-184. Routledge.
- Münster, Sander, Ferdinand Maiwald, Isabella di Lenardo, Juha Henriksson, Antoine Isaac, Manuela Milica Graf, Clemens Beck, and Johan Oomen. 2024. "Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe." *Heritage* 7 (2):794-816.
- Net, Francesc, Marc Folia, Pep Casals, and Lluís Gómez. 2023. "Transductive Learning for Near-Duplicate Image Detection in Scanned Photo Collections." *Document Analysis and Recognition - ICDAR 2023*, Cham, 2023//.
- Nofre, David, Mark Priestley, and Gerard Alberts. 2014. "When Technology Became Language: The Origins of the Linguistic Conception of Computer Programming, 1950–1960." *Technology and Culture* 55 (1):40-75.
- OCDE, Organización para la Cooperación y el Desarrollo Económico. s.d. Recomendación del consejo sobre gobierno abierto. OCDE.
- Omigie, Christopher Agbeniaru, Dorcas Ejemeh Krubu, and Oboite Solomon Anthony. 2023. "Exploring artificial intelligence for records and archival management system (AI-RAMS) for Nigerian public organizations." *International Journal of Library Science & Education Research* 29 (8):125-134.
- Oxman, Andrew D., and Gordon H. Guyatt. 1993. "The Science of Reviewing Research." *Annals of the New York Academy of Sciences* 703 (1):125-133. doi: doi.org/10.1111/j.1749-6632.1993.tb26342.x.
- Palomar, Núria, and Joan Pérez-Ventayol. 2023. "El govern de la dada a la Generalitat de Catalunya: instruments normatius i formes d'implantació." *Lligall* (46):118-127.
- Pasquale, Frank. 2015. *The black box society: The secret algorithms that control money and information*. Cambridge (MSS), EUA: Harvard University Press.
- Power, Daniel J. 2016. "'Big Brother' can watch us." *Journal of Decision Systems* 25 (sup1):578-588. doi: 10.1080/12460125.2016.1187420.
- Proctor, J., and R. Marciano. 2021. "An AI-Assisted Framework for Rapid Conversion of Descriptive Photo Metadata into Linked Data." 2021 IEEE International Conference on Big Data (Big Data), 15-18 Dec. 2021.

- Qian Sun, Tara, and Rony Medaglia. 2019. "Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare." *Government Information Quarterly* 36 (2):368-383. doi: doi.org/10.1016/j.giq.2018.09.008.
- Quijano-Sánchez, Lara, Federico Liberatore, José Camacho-Collados, and Miguel Camacho-Collados. 2018. "Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police." *Knowledge-Based Systems* 149:155-168. doi: doi.org/10.1016/j.knosys.2018.03.010.
- Ramió, Carles. 2019. *Inteligencia artificial y administración pública: robots y humanos compartiendo el servicio público*. 1a. ed. Madrid: Los libros de la Catarata.
- Raventós, Pepita, Celio Hernández, and Meritxell Simon. 2023. "IA y archivo : La tecnología de reconocimiento de textos manuscritos en fondos patrimoniales: un ejemplo de 10 diarios en la formación de maestras y maestros en el año 1932." *Tábula* (25):83-98. doi: 10.51598/tab.937.
- Reutter, L., and H. S. Spiíker. 2019. "The quest for workable data: Building machine learning algorithms from public sector archives." In *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms*, edited by Anna Tuschling, Andreas Sudmann and Bernhard J. Dotzler, 95-107. Transcript-Verlag, Bielefeld.
- Rhoads, J. 1969. "The historian and the new technology." *The American Archivist* 32 (3):209-213.
- Rolan, G., G. Humphries, L. Jeffrey, E. Samaras, T. Antsoupova, and K. Stuart. 2019. "More human than human? Artificial intelligence in the archive." *Archives and Manuscripts* 47 (2):179-203. doi: 10.1080/01576895.2018.1502088.
- Salvador Serna, Miquel. 2021. "Inteligencia artificial y gobernanza de datos en las administraciones públicas: reflexiones y evidencias para su desarrollo." *Gestión y Análisis de Políticas Públicas* (26):20-32.
- Samoili, Sofia, Montserrat Lopez Cobo, Blagoj Delipetrev, Fernando Martinez-Plumed, Emilia Gomez, and Giuditta De Prato. 2021. *AI Watch. Defining Artificial Intelligence 2.0. Towards an operational definition and taxonomy for the AI landscape*. edited by JRC Research Reports and European Comission. Faculty of Computer Science (Sevilla): Joint Research Centre (JRC).
- Serra Serra, Jordi. 2024. "El rol de la sèrie documental en entorns de gestió dadificats: una proposta metodològica en el marc de la governança de la informació." Research thesis, Biblioteconomia. Documentació (Ciències de l'Educació), Universitat de Barcelona
- Services, IBM Global. s.d. "¿Qué es la automatización inteligente?", accessed 14/03/2024. <https://www.ibm.com/es-es/topics/intelligent-automation>.
- Shabou, Basma Makhoulouf, Julien Tièche, Julien Knafo, and Arnaud Gaudinat. 2020. "Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data." *Records management journal* 30 (2):175-200.
- Sixto, A, R Aleixandre, A Vidal, R Lucas, and L Castelló. 2019. "Data sharing: qué son y cómo se pueden compartir los datos de investigación." In *Manual de*

recomendación para gestores de la información. SEDIC. Sociedad Española de Documentación e Información Científica España.

Smallwood, Robert F. 2019. *Information governance: Concepts, strategies and best practices*: John Wiley & Sons.

Soler Jiménez, J. 2020. "Governança de dades, gestió documental o governança de la informació?" *Escola d'Administració Pública de Catalunya*. <https://eapc.blog.gencat.cat/2020/09/29/governanca-de-dades-gestio-documental-o-governanca-de-la-informacio-joan-soler-jimenez/>.

Stielow, Frederick J. 1991. "Archival Theory Redux and Redeemed: Definition and Context toward a General Theory." *The American Archivist* 54 (1):14-26.

Strigel Carter, Kirsten, Abby Gondek, William Underwood, Teddy Randby, and Richard Marciano. 2022. "Using AI and ML to optimize information discovery in under-utilized, Holocaust-related records." *AI SOCIETY* 37 (3):837-858.

Stylianou, N., D. Vlachava, I. Konstantinidis, N. Bassiliades, and V. Peristeras. 2022. "Doc2KG: Transforming Document Repositories to Knowledge Graphs." *International Journal on Semantic Web and Information Systems* 18 (1):20. doi: 10.4018/ijswis.295552.

Tangi, Luca, Marco Combetto, Jaume Martin Bosch, and Paula Rodriguez Müller. 2023. *Artificial Intelligence for Interoperability in the European Public Sector*. Joint Research Centre (Seville).

Tarride, Solène, Mélodie Boillet, Jean-François Moufflet, and Christopher Kermorvant. 2023. "Simara: a database for key-value information extraction from full pages."

Teel, Zoë Abbie. 2024. "Artificial Intelligence's Role in Digitally Preserving Historic Archives." *Preservation, Digital Technology & Culture* 53. doi: 10.1515/pdte-2023-0050.

Theimer, Kate. 2018. "It's the end of the archival profession as we know it, and I feel fine." *Archival futures*:1-17.

Torreblanca Valverde, Sònia. 2021. "La perspectiva arxivística i de gestió documental en l'avaluació de les dimensions de qualitat de les dades obertes: una proposta teòrica i metodològica." Master thesis, Escola Superior d'Arxivística i Gestió de Documents, Universitat Autònoma de Barcelona.

Trehan, V. 2023. "AI-Powered Archives: Revolutionizing Information Access for the Future." 2023 IEEE International Conference on Big Data (BigData), 15-18 Dec. 2023.

Tricco, Andrea C., Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D. J. Peters, Tanya Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp, and Sharon E. Straus. 2018. "PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation." *Annals of Internal Medicine* 169 (7):467-473. doi: 10.7326/M18-0850.

- Tsabedze, Vusi. 2024. "Managing Records in the Age of Artificial Intelligence: How Prepared Are Archives and Records Management Professionals in Eswatini?" *Internet Reference Services Quarterly* 28 (1):77-95. doi: 10.1080/10875301.2023.2284898.
- Unió Europea, Comissió. 2021. "European Commission: White paper on artificial intelligence—a European approach to excellence and trust."
- Unió Europea, Comissió. s.d. "AI Act." accessed 15/03/2023. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.
- Valdiviezo Abad, Cesibel, and Tiziano Bonini. 2019. "Automatización inteligente en la gestión de la comunicación." *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales* (29):169-196. doi: 10.31921/doxacom.n29a9.
- Valero Torrijos, Julián. 2019. "Les garanties jurídiques de la intel·ligència artificial en l'activitat administrativa des de la perspectiva de la bona administració." *Revista catalana de dret públic* (58):82-96.
- Valle-Cruz, David, J. Ignacio Criado, Rodrigo Sandoval-Almazán, and Edgar A. Ruvalcaba-Gomez. 2020. "Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation." *Government Information Quarterly* 37 (4):101509. doi: doi.org/10.1016/j.giq.2020.101509.
- Van der Werf, Titia, and Bram Van der Werf. 2022. "Will archivists use AI to enhance or to dumb down our societal memory?" *AI & SOCIETY* 37 (3):985-988. doi: 10.1007/s00146-021-01359-x.
- Venkata, S. K., P. Young, M. Bell, and A. Green. 2022. "Alexa, Is This a Historical Record?" *Journal on Computing and Cultural Heritage* 15 (1):20. doi: 10.1145/3479008.
- Venkata, S. K., P. Young, and A. Green. 2021. "Text File Format Identification: An Application of AI for the Curation of Digital Records." 17th International Conference on Digital Preservation (iPRES 2021), Beijing, China, October 19-22, 2021.
- Vidal Ruiz, Enrique, Carlos Alonso Villalobos, Verónica Romero Gómez, Vicent Bosch Campos, Lourdes Márquez Carmona, María del Carmen Orcero, David Garrido Romero, and Mili Jiménez Melero. 2020. "Proyecto Carabela: un método revolucionario para la investigación de naufragios en archivos históricos basado en la inteligencia artificial." *Drassana: revista del Museu Marítim* (28):88-105.
- Vinuesa, Ricardo, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 2020. "The role of artificial intelligence in achieving the Sustainable Development Goals." *Nature Communications* 11 (1):233. doi: 10.1038/s41467-019-14108-y.
- Wang, H. 2022. "Providing public access to archival materials and the problem of fact-checking in the post-truth era." *Social Science Quarterly* 103 (7):1750-1764. doi: 10.1111/ssqu.13232.

Wilkinson, Mark D., Michel Dumontier, Isbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (1):160018. doi: 10.1038/sdata.2016.18.

Yang, H., J. Yang, J. Yang, J. Cai, and J. Wang. 2023. "Research on Application of Artificial Intelligence Technology in Document Management System." Lecture Notes in Electrical Engineering.

ÍNDIX DE FIGURES

Fig. 1 Selecció, avaluació i inclusió dels estudis (presentats mitjançant el diagrama de flux PRISMA)	30
Fig. 2 Gràfic lineal que mostra els anys de publicació dels estudis seleccionats	36
Fig. 3 Bases de dades d'obtenció dels estudis seleccionats	37
Fig. 4 Mapa coroplètic que mostra els països d'aplicació dels estudis seleccionats, i la quantitat d'estudis duts a terme	41
Fig. 5 Mètodes d'investigació utilitzats en els estudis seleccionats per la revisió	42
Fig. 6 Enfocament metodològic utilitzat en els estudis seleccionats	43
Fig. 7 Nombre de citacions en les principals bases de dades utilitzades	45
Fig. 8 Tipus d'IA analitzada o aplicada als estudis seleccionats (agrupades per tipus i especificant les tècniques utilitzades)	62

ÍNDIX DE TAULES

Taula 1. Definició de conceptes entorn la IA. Fonts: (Samoili et al. 2021, Jarvinen 2020, Frankish and Ramsey 2014)	12
Taula 2. Vocabulari utilitzat en la primera cerca. Obtingut de la base de dades terminològica del projecte InterPARES AI	26
Taula 3. Estratègies de cerca realitzades en cada repositori/base de dades amb els connectors booleans i paraules clau (ítems: <i>database / search strategy</i>)	29
Taula 4. Categories d'anàlisi que han permès la recopilació d'informació per cada estudi seleccionat	32
Taula 5. Informació bàsica a nivell descriptiu dels estudis seleccionats	35

Taula 6. Afiliacions dels/les autors/es més productius/es segons el nombre d'estudis realitzats i àrees de coneixement més representades (#pub: nombre de publicacions, #autors: nombre d'autors)	38
Taula 7. Resum dels objectius dels estudis seleccionats per la revisió sistemàtica	41
Taula 8. Beneficis de l'ús de la IA en els arxius públics a partir dels resultats dels estudis seleccionats.....	48
Taula 9. Riscos de l'ús de la IA en els arxius públics a partir dels resultats dels estudis seleccionats	52
Taula 10. Requisits en l'ús de la IA als arxius públics, a partir dels resultats dels estudis seleccionats.....	55
Taula 11. Reptes de futur en l'ús de la IA als arxius públics, a partir dels resultats dels estudis seleccionats.....	59