

L'intelligenza artificiale al servizio degli archivi digitali: una survey sulle applicazioni per la ricostruzione delle aggregazioni documentali e l'arricchimento dei metadati

Stefano Allegrezza

Università degli Studi di Macerata (Italia)
stefano.allegrezza@unimc.it.

Abstract

Questo contributo intende presentare i risultati di uno studio condotto nell'ambito del progetto internazionale InterPARES Trust AI ed intitolato «The role of AI in identifying or reconstituting archival aggregations of digital records and enriching metadata schemas». L'obiettivo generale dello studio è quello di indagare la capacità dell'intelligenza artificiale di aiutare nelle operazioni di creazione (o ri-creazione) di aggregazioni archivistiche allo scopo di risolvere il problema della presenza di documenti non aggregati, non ordinati o de-contestualizzati (sia nella fase corrente che in quella semi-corrente dell'archivio) che si presenta in molte situazioni. Spesso, infatti, sia nelle amministrazioni pubbliche che nelle aziende, i documenti non vengono né classificati né fascicolati; in altri casi, le aggregazioni documentali vengono formate ma in modo non corretto. Inoltre, non di rado i metadati – che sono necessari per garantire l'autenticità, l'affidabilità e la ricercabilità – non vengono correttamente individuati ed associati ai documenti. In questo modo l'archivio dell'organizzazione non viene formato correttamente, e ciò costituisce una grave criticità perché fa sì che ci sia un numero incontrollato di documenti non ordinati, mal collocati e quindi difficili da trovare. Purtroppo, nonostante i progressi compiuti dalle tecnologie informatiche per fornire un aiuto nelle attività di gestione documentale, bisogna riconoscere che gli attuali prodotti software sono in grado di fornire un supporto ancora molto limitato a questo tipo di esigenze. Tuttavia, le tecniche di intelligenza artificiale sembrano promettere grossi passi avanti in questo campo. Lo studio che viene presentato con questo contributo si è posto l'obiettivo di affrontare tali problemi fornendo una risposta a domande di ricerca come le seguenti: gli strumenti di intelligenza artificiale possono aiutare a creare le aggregazioni documentali quando queste non sono mai state formate o a ri-crearle quando erano state formate ma sono andate perdute? Possono fornire un valido aiuto nell'individuazione di metadati e nella associazione ai documenti relativi?

Parole chiave: Intelligenza artificiale, archivi, fascicoli, metadati, aggregazioni documentali, vincolo archivistico, AIUCD2024

This paper aims to present the results of a study conducted as part of the international InterPARES Trust AI project and entitled «The role of AI in identifying or reconstituting archival aggregations of digital records and

enriching metadata schemas». The overall objective of this study is to investigate the ability of artificial intelligence to support the creation (or re-creation) of archival aggregations to solve the problem of the presence of unaggregated, unordered or de-contextualised documents (both in the current and semi-current phase of the archive) that arises in many situations. Often, in both public administrations and companies, documents are neither classified nor aggregated; or document aggregations are formed but incorrectly. Furthermore, metadata – which are necessary to guarantee authenticity, reliability and searchability – are not often correctly identified and associated with documents. As a result, the organisation’s archive is not properly formed, which is a severe problem because it leads to an uncontrolled number of documents that are unordered, misplaced and therefore difficult to find. Unfortunately, in spite of the progress made by information technology to provide help in document management activities, it must be recognised that current software products are able to provide very limited support for this type of need. However, artificial intelligence techniques seem to promise great progress in this field. The study aims to address these issues by providing answers to research questions such as the following: can artificial intelligence tools help to create document aggregations when these have never been formed or to re-create them when they had been formed but have been lost? Can they provide valuable help in identifying metadata and associating it with related documents?

Keywords: Artificial intelligence, archives, case-files, metadata, document aggregations, archival bond, AIUCD2024

1. Introduzione

La formazione delle aggregazioni documentali¹ rientra tra le operazioni archivistiche strategiche ai fini della corretta formazione dell’archivio. Infatti «la decisione di aggregare un nuovo documento archivistico ad un fascicolo già aperto oppure di aprire un nuovo fascicolo in seguito all’acquisizione di un certo documento archivistico consente l’ordinato stratificarsi della produzione documentaria di un ente nel corso della sua concreta e quotidiana attività amministrativa all’interno di un quadro logico astratto, cioè il titolario di classificazione» ([4], 80). Tale operazione è fondamentale anche nel caso degli archivi nativi digitali, tanto che, con riferimento al contesto italiano, l’articolo 64, comma 4, del «Testo Unico sulla documentazione amministrativa» prevede che «le amministrazioni determinano autonomamente e in modo coordinato per le aree organizzative omogenee, le modalità di attribuzione dei documenti ai fascicoli che li contengono e ai relativi procedimenti, definendo adeguati piani di classificazione d’archivio per tutti i documenti, compresi quelli non soggetti a registrazione di protocollo» ([8]). Tale previsione viene ripresa anche dall’art. 41 (“Procedimento e fascicolo informatico”) del “Codice dell’amministrazione digitale” il quale stabilisce che «la pubblica amministrazione titolare del procedimento raccoglie in un fascicolo informatico gli atti, i documenti e i dati del procedimento medesimo da chiunque formati» ([9]). Da ultimo, anche le “Linee Guida sulla formazione, gestione e conservazione dei documenti informatici” emanate dall’Agenzia per l’Italia Digitale (AgID) stabiliscono che le Pubbliche Amministrazioni gestiscono «i flussi

¹ Si ricorda che, ai sensi delle “Linee guida sulla formazione, gestione e conservazione di documenti informatici” pubblicate dall’Agenzia per l’Italia Digitale (AgID) [1], sono previste tre tipologie di “aggregazioni documentali”: 1) il *fascicolo*, aggregazione documentale informatica strutturata e univocamente identificata contenente atti, documenti o dati informatici prodotti e funzionali all’esercizio di una attività o allo svolgimento di uno specifico procedimento; 2) le *serie documentali*, costituite da documenti singoli accorpati per ragioni funzionali in base alla tipologia di riferimento; 3) le *serie di fascicoli*, costituite da fascicoli accorpati per ragioni funzionali in base alla classe di riferimento o alla tipologia di fascicoli.

documentali mediante fascicoli informatici predisposti secondo il piano di classificazione e relativo piano di organizzazione delle aggregazioni documentali» ([1]).

La creazione dei fascicoli, quindi, non è soltanto una buona prassi dal punto di vista archivistico, ma è anche un obbligo normativo. Purtroppo, la situazione che si riscontra diffusamente, sia nelle realtà pubbliche che in quelle private, è una insufficiente attenzione nei confronti di questa operazione, tanto necessaria – in quanto la sua mancanza porta alla formazione di archivi disordinati, caotici e molto difficili da riordinare in una fase successiva – quanto trascurata. In molte amministrazioni pubbliche i documenti non sono classificati né aggregati in fascicoli. In altri casi, le aggregazioni documentali sono create ma non correttamente, con il risultato della presenza di un numero incontrollato di documenti non ordinati, non collocati nel fascicolo corretto e difficili da trovare. In molti casi mancano i metadati, «utilizzati per finalità molteplici: ricerca, gestione e localizzazione, selezione, interoperabilità» ([12], 189) e necessari per garantire l'affidabilità, l'autenticità, la qualità.

Nonostante i progressi compiuti dalle varie tecnologie a supporto della gestione dei documenti, la disponibilità di software per questo tipo di attività rimane ad oggi ancora molto limitata. Il motivo per cui l'operazione di formazione delle aggregazioni documentali è così trascurata è riconducibile a diverse motivazioni, come la mancanza di conoscenze archivistiche, la poca consapevolezza della sua importanza, la mancanza di volontà, o, infine, la mancanza di tempo. Tutto ciò ha delle conseguenze rilevanti non solo in termini di perdita di efficienza dell'azione amministrativa, causata dalla difficoltà di gestire il flusso procedimentale dal momento che i documenti ad esso relativi non si trovano riuniti all'interno della stessa aggregazione documentale, ma anche dal punto di vista della preservazione della memoria, dal momento che un archivio del genere sarà anche difficilmente fruibile da parte delle generazioni future di ricercatori e studiosi. Fino ad oggi la soluzione a questo problema è stata cercata soprattutto nell'organizzazione di interventi mirati di formazione con lo scopo di sensibilizzare gli operatori degli uffici sull'importanza dell'operazione di fascicolazione, ma questi interventi non sempre si sono rivelati efficaci e spesso non hanno sortito l'effetto desiderato.

Viene naturale, allora, chiedersi se le tecnologie di intelligenza artificiale possano essere utili per operazioni quali la classificazione dei documenti, la creazione di aggregazioni documentali o l'arricchimento dei metadati, svolgendo così – in maniera automatica o semi-automatica – quei compiti che l'essere umano non riesce a compiere per i motivi più disparati.

Un caso paradigmatico è quello relativo alla gestione dell'archivio di posta elettronica, che è una delle attività più dispendiose in termini di tempo sia nel settore pubblico che in quello privato. La dottrina archivistica insegna che le email ricevute ed inviate, così come in generale tutta la documentazione, andrebbero ordinate all'interno di una struttura articolata in fascicoli e sotto-fascicoli ('cartelle' e 'sotto-cartelle' nel linguaggio informatico), create in relazione ai vari 'affari' trattati, ai vari progetti, alle varie attività, oppure in relazione ai vari corrispondenti; l'articolazione di questa struttura «non dovrebbe essere né eccessivamente dettagliata (perché altrimenti l'assegnazione di ciascuna email nella cartella di competenza richiederebbe troppo tempo) né eccessivamente generale (perché altrimenti perderebbe di efficacia); di solito un numero di livelli compreso tra due e tre è la scelta migliore» ([3], 196-197). Quello che andrebbe assolutamente evitato è di lasciare tutte le e-mail nelle cartelle della "Posta in arrivo" e della "Posta inviata" dal momento che così facendo ben difficilmente sarà possibile mettere ordine in quella massa disordinata e confusa di e-mail, che spesso raggiunge dimensioni importanti (nell'arco di una vita può facilmente raggiungere la dimensione di decine di migliaia di messaggi e a volte anche di più). Purtroppo, riuscire a formare un archivio organizzato in questo modo non è semplice dal momento che aggregare nelle relative cartelle – del client di posta o dell'applicazione webmail –

la quantità spesso eccessiva di e-mail che quotidianamente vengono ricevute ed inviate richiederebbe di dedicare un tempo non indifferente a questa attività. Questo è probabilmente il motivo per cui la maggior parte delle persone – salvo casi particolari – lascia tutte le e-mail nelle cartelle della “posta in arrivo” e della “posta inviata” creando così un archivio sostanzialmente privo di aggregazioni documentali dove rintracciare una e-mail è ben difficile (se non affidandosi alle funzioni di ricerca, che però richiedono che la e-mail ricercata abbia almeno un oggetto ben definito)² e dove nel futuro sarà molto difficile condurre studi e ricerche.

Anche in questo caso viene spontaneo chiedersi se via siano applicazioni basate sull’intelligenza artificiale in grado di riunire all’interno di una aggregazione documentale le e-mail scambiate con un certo corrispondente o relative ad un determinato ‘affare’, delegando, cioè, all’intelligenza artificiale lo svolgimento di quelle operazioni che l’uomo non riesce a svolgere.

2. Metodologia

Per ovviare alle criticità appena evidenziate, all’interno del progetto di ricerca internazionale “InterPares Trust AP” ([14]) al gruppo di lavoro denominato CU05 (“Creation and Use 05”) è stato affidato il compito di condurre uno studio per indagare se le tecniche di intelligenza artificiale possono essere d’aiuto nelle operazioni di classificazione dei documenti, di formazione delle aggregazioni documentali e di individuazione dei metadati. L’obiettivo era anche quello di comprendere se le tecniche di intelligenza artificiale sono in grado di ristabilire il legame archivistico tra una moltitudine di documenti decontestualizzati e di integrare ed arricchire schemi di metadati incompleti.

L’indagine ha avuto inizio con un censimento delle aziende che dichiarano di sviluppare prodotti basati sull’intelligenza artificiale applicata ai settori della gestione documentale e che sono stati giudicati rilevanti nell’ambito dello studio. Questa prima fase, che si è svolta tra febbraio e giugno 2022, ha portato all’identificazione di un elenco iniziale – non esaustivo – di circa 300 aziende. L’elenco è stato costruito sia mediante ricerche dirette svolte su Internet utilizzando parole chiave³ e stringhe di testo, sia grazie alle risorse e conoscenze messe a disposizione da alcuni professionisti [cfr. sez. “Ringraziamenti” alla fine del contributo]. Le caratteristiche dei software di intelligenza artificiale sono state analizzate in prima battuta in base alle informazioni

² Si potrebbe obiettare che le funzioni di ricerca consentono di effettuare ricerche di tipo *full-text* – che prendono in considerazione anche il corpo del messaggio – e, combinate con i filtri di ricerca (mittente, destinatario, data, presenza di allegati, ecc.) potrebbero consentire di trovare una determinata e-mail. Su questo punto occorre fare un paio di considerazioni. Innanzitutto, le ricerche di questo tipo non sempre consentono di individuare l’e-mail che interessa perché spesso restituiscono un numero molto elevato di risultati, specialmente quando i parametri di ricerca non sono ben definiti (ad esempio, non ci si ricorda che è il mittente, quale fosse l’oggetto dell’e-mail, o il periodo in cui è stata ricevuta o inviata, ecc.). In secondo luogo, quando si parla di ricerca archivistica non è sufficiente il ritrovamento di una singola e-mail ma occorre ritrovare l’intero ‘carteggio’ elettronico intercorso con un certo corrispondente o relativamente ad un certo ‘affare’: questo tipo di risultato lo si ottiene solo con una corretta organizzazione dell’archivio di posta elettronica.

³ Ad esempio, sono state effettuate ricerche utilizzando stringhe di ricerca come “artificial intelligence”, “records”, “documents”, “information extraction”, “archival bond”; oppure “artificial intelligence”, “records”, “document classification”, “file plan”, “archives”; o, ancora, “artificial intelligence”, “records management”, “document classification”, “recordkeeping”, “archives”. La ricerca è stata condotta anche utilizzando altre lingue oltre a quella inglese.

disponibili sui rispettivi siti web (ad esempio, se l'azienda dichiara che la propria applicazione di intelligenza artificiale è applicabile al settore della gestione dei documenti o a quello della conservazione degli archivi, anche se in alcuni casi non vi è una dichiarazione esplicita, ma la si può solo intuire dal contenuto del sito web).

Successivamente, dall'elenco iniziale è stato selezionato un elenco ristretto di 100 aziende – la cui distribuzione geografica è riportata in Figura 1 – identificate come potenzialmente molto interessanti per gli scopi dello studio e che sarebbero state ulteriormente vagliate. Infine, l'elenco delle 100 aziende è stato ulteriormente ridotto andando a selezionare quelle aziende che – sulla base del portafoglio clienti, del coinvolgimento diretto nel settore della gestione documentale, della conformità ai quadri normativi e agli standard relativi alla gestione degli archivi e dei documenti e agli standard di settore e della reputazione generale dell'azienda – meglio rispondevano agli scopi dello studio. Da questa ulteriore fase di 'scrematura', che si è svolta tra il giugno e l'agosto 2022, è emerso un elenco di 28 aziende – potenzialmente le più interessanti, ma evidentemente l'elenco potrebbe non essere esente da errori e fraintendimenti – alle quali è stata inviata una lettera ufficiale (in inglese, spagnolo o portoghese, a seconda della lingua preferita dall'azienda) con l'invito a partecipare all'indagine.



Figura 1. Distribuzione geografica delle 100 aziende selezionate.

Allo scopo di raccogliere in maniera sistematica le informazioni sugli applicativi di intelligenza artificiale fornite dalle aziende intervistate, il gruppo di lavoro ha preparato un dettagliato questionario in lingua inglese, necessario per un'adeguata valutazione delle applicazioni destinate a supportare la ricostituzione delle aggregazioni archivistiche e l'arricchimento dei metadati. Il questionario è composto da quattro sezioni, ciascuna con diverse domande aperte.

La sezione I (*Risultati conseguiti*) si concentra sui risultati conseguiti dalle aziende, in particolare sulle caratteristiche delle applicazioni, sulle piattaforme di sviluppo, sulle caratteristiche principali e sui punti di forza, sugli aspetti da migliorare, sugli sviluppi futuri, nonché, infine, sulla conformità agli standard archivistici e di gestione dei documenti. La sezione II (*Capacità specifiche*)

per i sistemi di gestione documentale e di posta elettronica) si occupa delle funzionalità specifiche delle applicazioni per la gestione dei documenti, compresa quelli dei sistemi di posta elettronica. Le domande riguardano l'automazione di diverse attività di gestione dei documenti, come l'archiviazione dei documenti in fascicoli secondo uno schema di classificazione dei documenti, la valutazione e lo scarto dei documenti secondo un piano di conservazione, l'estrazione di metadati e l'indicizzazione dei documenti per fornire informazioni su aggregazioni correlate. La sezione III (*Tecnologie e metodi utilizzati nelle applicazioni di intelligenza artificiale*) analizza le tecnologie e i modelli utilizzati, le strategie e le tecniche di apprendimento automatico, nonché gli elementi dei documenti o dei metadati che l'applicazione prende in considerazione per prendere decisioni e fare inferenze. Infine, la sezione IV (*Controlli di audit – indicatori di prestazione*) si concentra sulle verifiche e sugli indicatori di prestazione per misurare la percentuale di successo e di insuccesso delle applicazioni e le eventuali distorsioni.

Alla lettera di invito hanno risposto positivamente 13 aziende (vd. Figura 2) che hanno accettato di partecipare all'indagine e con ciascuna delle quali è stato organizzato un incontro on-line per illustrare il questionario e le sue finalità. All'incontro, oltre al gruppo di lavoro, hanno partecipato i referenti delle aziende intervistate: di solito hanno partecipato coloro che si occupavano della gestione documentale, gli ingegneri informatici e gli archivisti (se presenti); si è privilegiata la partecipazione del personale tecnico e si è cercato di evitare la partecipazione dei responsabili commerciali (che avrebbero probabilmente esaltato le caratteristiche delle loro soluzioni fornendo così informazioni distorte). Successivamente, le aziende hanno compilato il questionario che è stato reso disponibile su Google Forms fino al mese di febbraio 2023.



Figura 2. Le 13 aziende che hanno risposto al questionario.

3. Risultati ottenuti

Il gruppo di lavoro ha preso in esame le risposte ottenute e ha elaborato un report che è stato pubblicato nel mese di novembre 2023 sul sito del progetto internazionale ([15]). Non essendo possibile, per mancanza di spazio, discutere in questa sede tutte le risposte ottenute nel questionario, si riportano alcune conclusioni di carattere generale.

Innanzitutto, l'indagine conteneva alcune domande sulle soluzioni tecniche adottate dalle aziende, e più precisamente sui modelli di analisi e sui tipi di tecniche utilizzate nei prodotti, sulle strategie di formazione, sugli elementi informativi elaborati dalle piattaforme e sulle caratteristiche degli ecosistemi informatici di cui i prodotti basati sull'IA hanno bisogno per funzionare correttamente. Tutte le aziende intervistate hanno sviluppato soluzioni basate su tecnologie di intelligenza artificiale per l'indicizzazione, la categorizzazione o la classificazione di documenti strutturati, semi-strutturati e non strutturati, sulla base di tecniche di apprendimento automatico e di estrazione automatica dei dati. Per quanto riguarda i modelli di analisi, il panorama risulta molto variegato: le aziende hanno dichiarato di utilizzare complessivamente 24 modelli diversi. *Neural networks* (7 volte) e *support vector machines* (4 volte) sono i modelli più citati (va notato che in alcuni casi le risposte sono state piuttosto generiche). Per quanto riguarda le tecniche utilizzate, sono state citate complessivamente ben 18 diverse tipologie (vd. Figura 3); la *classification* (9 volte) e il *clustering* (5 volte) sono state le risposte più comuni (questo era prevedibile, dato che le aziende sono state selezionate per la loro esperienza nella gestione dei documenti).

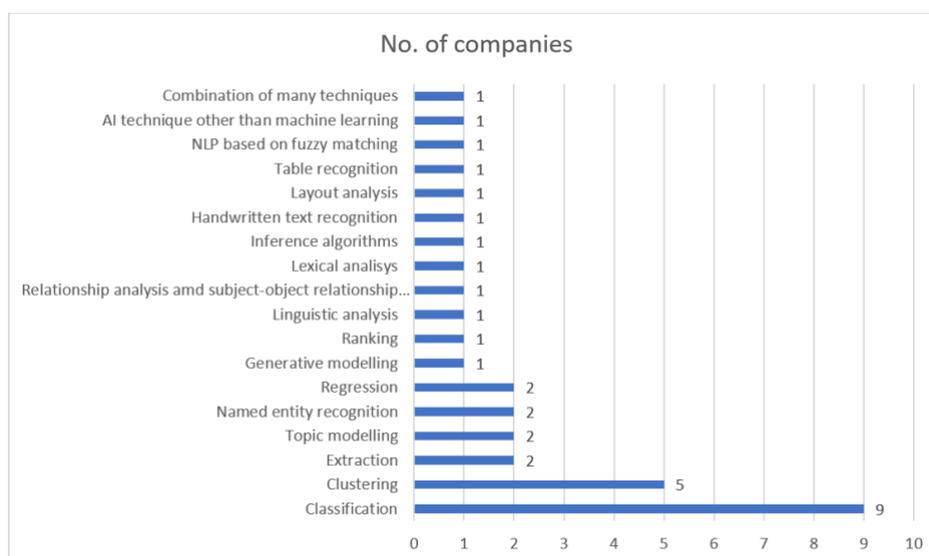


Figura 3. Le tecniche di intelligenza artificiale dichiarate dalle aziende che hanno risposto al questionario.

Infine, per quanto riguarda le strategie di apprendimento per i prodotti basati sull'IA, l'indagine ha mostrato una situazione mista (vd. Figura 4): 11 aziende utilizzano l'apprendimento *supervised*, 6 aziende l'apprendimento *unsupervised*, 4 aziende l'apprendimento *semi-supervised*, 2 aziende l'apprendimento *auto-supervised* e 2 aziende l'apprendimento *rule-based*. Naturalmente, diverse aziende utilizzano più di un tipo di strategie di apprendimento. Relativamente alle caratteristiche degli ecosistemi informatici necessari per far funzionare i prodotti basati sull'IA, la maggior parte dei prodotti che sono stati esaminati possono funzionare in ogni ambiente operativo (Windows, Linux, Mac, ecc.) e interagire con un numero molto elevato di piattaforme e applicazioni grazie ad API e connettori personalizzati, il cui sviluppo richiede, tuttavia, investimenti in termini di risorse finanziarie ed umane. Infine, quasi tutti i prodotti basati sull'intelligenza artificiale recensiti possono essere distribuiti e funzionare sia in locale che su cloud.

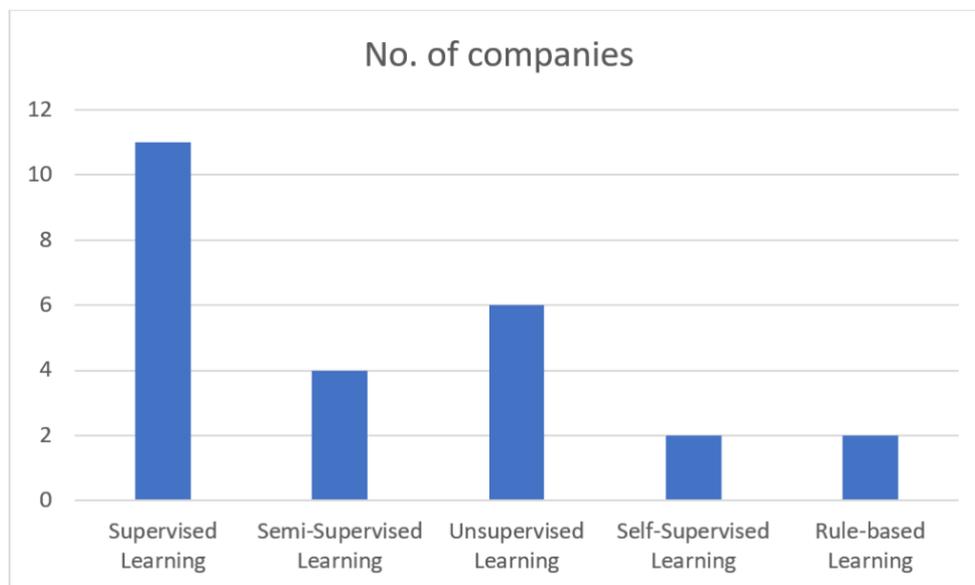


Figura 4. Le strategie di apprendimento dichiarate dalle aziende che hanno risposto al questionario

Per quanto riguarda la possibilità di creazione o ricostituzione di aggregazioni documentali basata su tecniche di intelligenza artificiale, le risposte ottenute non sono molto incoraggianti, in quanto è emerso come questa possibilità sia limitata a casi molto specifici: ad esempio, solo per determinate tipologie di documenti, oppure quando vi sono almeno delle informazioni di base sulla struttura delle aggregazioni, o, ancora, quando l'aggregazione automatica viene solamente proposta e necessita di una fase di convalida da parte dell'utente.

3.1 La classificazione automatica

Attraverso la classificazione i documenti vengono organizzati «in raggruppamenti di natura logica secondo una struttura di voci gerarchicamente organizzata (categorie, classi, sottoclassi) che rappresentano in modo sistematico le funzioni e le materie attribuite al soggetto produttore – nel caso di una persona giuridica – da provvedimenti istitutivi e regolamentari» ([6], 209).

La funzione di classificazione automatica dei documenti sulla base di uno schema di classificazione è offerta da quasi tutte le aziende che hanno risposto ai questionari (10 su 13, vd. Figura 5), comprese quelle che non si occupano specificamente di archivi e gestione dei documenti. Tra le aziende le cui piattaforme non prevedono la classificazione automatica, una si occupa principalmente del riconoscimento di testi scritti a mano e le altre due si concentrano sull'indicizzazione e sull'estrazione di elementi di metadati. Alcune di queste affermano che le loro applicazioni potrebbero comunque essere addestrate per classificare i documenti.

Tre aziende con competenze specifiche in materia di gestione documentale hanno risposto che le loro piattaforme sono in grado di analizzare i metadati disponibili sia nei documenti che nelle aggregazioni di cui fanno parte. In particolare, la prima ha dichiarato che la sua applicazione di intelligenza artificiale classifica i documenti «according to document type and case-folder specifications»; la seconda ha affermato di essere in grado di classificare i documenti e le aggregazioni «based on their content and context, against function-based records classification

schemes»; la piattaforma sviluppata dalla terza cerca inizialmente di classificare i documenti e le aggregazioni analizzando «any metadata fields that are found or inferred on records». Nel caso in cui i metadati disponibili si rivelino insufficienti per la classificazione, l'applicazione «will use a machine learning model (if configured) to classify the record based on its text». L'azienda ha, inoltre, sottolineato che «taxonomy, rules and ML model are fully customizable by each customer and can be used to reflect any classification scheme and administrative processes required», anche se, per quanto riguarda i processi amministrativi, al momento «only disposal workflows are currently supported based on the classification».

Altre sei aziende (due delle quali hanno competenze nel settore della gestione dei documenti) hanno affermato che le loro piattaforme sono in grado di classificare i documenti sulla base di uno schema di classificazione dei documenti: una di esse ha anche aggiunto di aver sviluppato «a classifier that is able to classify documents and images».

L'ultima azienda ha dichiarato che la sua piattaforma può essere addestrata a riconoscere i tipi di documenti e ad applicare qualsiasi tipo di schema di classificazione; più precisamente, la piattaforma è capace di «process files and records (with suitable connectors to the source systems) and generate labels and tags belonging to any record classification scheme (taxonomy or term ontology)». In accordo con l'approccio generale adottato da questa azienda per sviluppare i suoi prodotti, la piattaforma di intelligenza artificiale può essere addestrata dagli stessi utenti, che possono alimentarla con specifici set di dati: sulla base dei dati in input, la piattaforma costruisce un grafo di conoscenza che verrà progressivamente affinato e utilizzato per classificare documenti e aggregazioni.



Figura 5. Le aziende che hanno dichiarato capacità di classificazione

3.2 La ricostruzione delle aggregazioni documentali

La creazione delle aggregazioni documentali è l'operazione attraverso cui si accorpano i documenti, «fisicamente se cartacei, logicamente se elettronici, in modo da soddisfare le esigenze

di lavoro» ([6]:211). Nella tradizione archivistica, si parla di “formazione dei fascicoli”, “fascicolazione” o “fascicolatura” ([17]), ma la normativa italiana prevede anche altre tipologie di aggregazioni documentali oltre al fascicolo⁴. Mentre la classificazione guida la sedimentazione dei documenti in modo ordinato e coerente, la fascicolazione aggrega tutti i documenti prodotti da una stessa attività o processo amministrativo in unità archivistiche ([10]; [13]:26). Pertanto, «classes and files are separate but interrelated entities of the same structure [the records classification scheme, integrated with the file plan]. Classes represent the functions and activities attributed to a records creator through regulation. They form an abstract structure in which, generally at the last classification level, files are created. Records are preferably placed into files or are logically linked to them» ([5]:38).

Secondo le risposte fornite al questionario, dieci delle tredici aziende intervistate (vd. Figura 6) hanno dichiarato che le proprie applicazioni sono in grado di archiviare i documenti nelle relative cartelle, fascicoli o aggregazioni in modo automatico o semi-automatico, anche se con delle limitazioni. Le motivazioni addotte sono state le più disparate: «there are some automatizations with the document type (associated with a folder or case-folder)»; «records are aggregated to a parent that users can specify per content source, and which is reflective of the structure of the content source (e.g. an email inbox, a folder on a file share)»; «it can be used to automatically (or, in some cases, semi-automatically, with a final user validation in the case of a human-in-the-loop workflow) generate metadata labels, tags and other information on the base of the document content». La creazione di etichette e tag può seguire «one or more classification or labelling schemes». Tra le tre aziende che non forniscono la fascicolazione automatica, una ha dichiarato che la sua applicazione «is capable of automatically categorising records based on their metadata and content rules created by users, but it does not file items to folders at this time», ma l'azienda sta lavorando per sviluppare questa funzione.

Ulteriori domande sono state poste in relazione alla creazione di aggregazioni, come la possibilità che le applicazioni facciano inferenze su quali documenti appartengono (o potrebbero appartenere) alla stessa aggregazione (ad esempio, lo stesso fascicolo per affare, lo stesso fascicolo per soggetto, la stessa serie, ecc.). Le risposte sono state positive in nove casi su tredici, e le aziende hanno dichiarato che vengono fatte delle inferenze «based on content and/or context»; o se si verificano condizioni e situazioni particolari: «case-folder might be found using IA models in a trained system»; «only with user-created content rules to specify the categories and requirements for categorization. Once these content rules are created and enabled, the system can make these decisions on newly added records»; «if there is metadata to represent those processes (e.g. a case file number)»; «the tools can output aggregated extractions exposing specific relations between extracted entities».

⁴ Secondo l'Allegato 5 (“Metadati”) alle “Linee guida sulla formazione, gestione e conservazione di documenti informatici” dell'Agenzia per l'Italia Digitale ([2]) si distinguono quattro tipologie di fascicoli: il fascicolo *per affare*, che conserva i documenti relativi a una competenza non proceduralizzata – ma che nella consuetudine amministrativa la pubblica amministrazione deve concretamente portare a buon fine – e ha una data di apertura e una durata circoscritta; il fascicolo *per persona fisica* o *giuridica*, che comprende tutti i documenti, anche con classifiche diverse, che si riferiscono a una persona (fisica o giuridica); il fascicolo *per attività*, che comprende i documenti prodotti nello svolgimento di un'attività amministrativa semplice che implica risposte obbligate o meri adempimenti, per la quale quindi non è prevista l'adozione di un provvedimento finale ed in genere ha una durata annuale; il fascicolo *per procedimento amministrativo*, che conserva una pluralità di documenti che rappresentano azioni amministrative omogenee e destinate a concludersi con un provvedimento amministrativo. I fascicoli sono cinque se si considerano distinti i fascicoli di persona fisica da quelli di persona giuridica. Cfr. ([1]:23) e ([2]:51). Cfr. anche [17].

Un'azienda ha risposto che la sua applicazione non è grado di fare inferenze direttamente, ma può fornire dati che possono essere utilizzati a questo scopo.

Quando alle aziende è stato chiesto se la propria applicazione di intelligenza artificiale fosse in grado di fare inferenze sull'organizzazione o sulla persona che ha prodotto i documenti, anche quando mancano metadati rilevanti per la loro identificazione, il numero di risposte positive è sceso a 7 su 13. Chi ha risposto positivamente ha dichiarato che la possibilità di ottenere dei risultati validi si basa sulla presenza di particolari situazioni: «getting the right case-folder and document type, you can establish sender and receiver of the record»; «if the person, or organisation is clearly marked on the document – e.g. with a stamp – this should be possible»; «If the involved entities are stated in the content of the document and the type of relation linking them is linguistically expressed, the technology is able to make such inferences»; «if the data somewhere exists in the records, then without metadata identification is possible»; «it will only identify organisations or persons referred to in the record text. It will not infer the role of that entity in the provenance of the record»; «Yes, based on content and/or context»; «possibly, but through classification».

Coloro che hanno risposto in modo negativo, hanno affermato che l'inferenza non può essere fatta direttamente senza metadati. Ad esempio: «our entity extraction feature can identify people, places, and things from the extracted text of a document, but if metadata is missing there is no framework as of yet to automatically make inferences about it».

In sintesi, le aziende che sviluppano applicazioni in grado di classificare automaticamente affermano che esse sono anche in grado di fascicolare i documenti o di collegare logicamente i documenti alle relative aggregazioni, nonché di fare inferenze per raggiungere questo obiettivo. Esse affermano, inoltre, che questa automazione può basarsi su schemi di classificazione o di etichettatura e dipende dal modo in cui le applicazioni funzionano o su come vengono addestrate, cioè in base al tipo di documento, al fascicolo, ai metadati, al contenuto del documento e al contesto.

Queste affermazioni sono forse un po' troppo ottimistiche. Sia la classificazione che la fascicolazione hanno connotazioni diverse a seconda delle varie tradizioni archivistiche o dei diversi standard di gestione dei documenti adottati. In alcuni casi le due operazioni sono nettamente distinte, in altri casi sono considerate sostanzialmente equivalenti e tendono a confondersi. Questo si riflette in qualche modo nelle risposte al questionario, dove non è sempre facile capire se le applicazioni categorizzano e/o archiviano i documenti in base a classi e/o fascicoli prestabiliti, e in che misura procedono sulla base di un piano di classificazione dei documenti e/o di un piano di fascicolazione, o anche solo basandosi su un elenco prestabilito di etichette o tag.

Va precisato che non tutte le aziende avevano familiarità con la terminologia archivistica utilizzata nel questionario, anche se il senso delle domande è stato spiegato durante le interviste introduttive. Questo potrebbe aver causato una mancata comprensione delle domande e alcune risposte non chiare. Per questo motivo, l'applicazione di queste piattaforme di intelligenza artificiale a specifiche esigenze di gestione dei documenti dovrà essere testata in modo approfondito nella seconda parte di questo progetto, che si concentrerà sulla sperimentazione delle applicazioni con corpora di documenti che saranno messi a disposizione come casi-studio da parte delle istituzioni partecipanti al progetto.

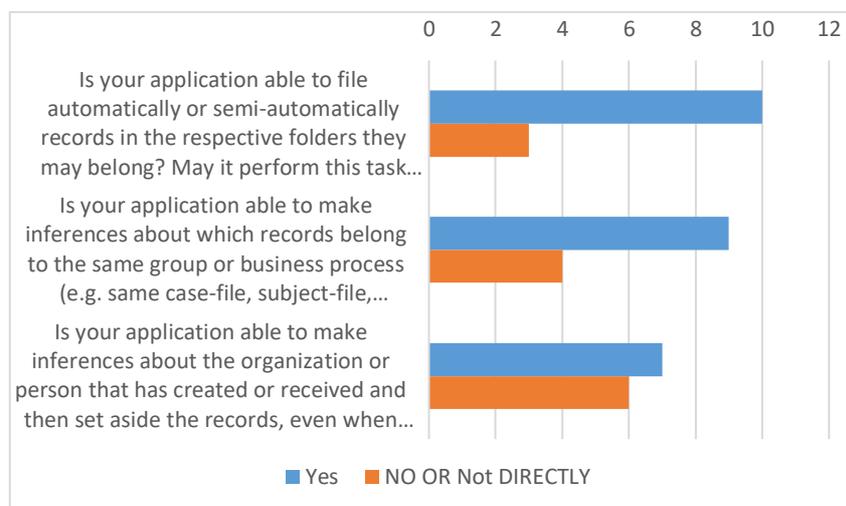


Figura 6. Le domande e relative risposte sulle capacità di aggregazione dei documenti

3.3 La ricostituzione del vincolo archivistico

Una parte del questionario era rivolta ad indagare se le tecnologie di intelligenza artificiale sono in grado di ristabilire il vincolo archivistico⁵ di documenti non aggregati, non ordinati o decontestualizzati, sia nella fase attiva (archivio corrente) che in quella semi-attiva (archivio di deposito) del loro ciclo di vita. Alle aziende è stato chiesto in particolare se le loro applicazioni sono capaci di ricostituire aggregazioni archivistiche andate perdute. Cinque su tredici (vd. Figura 7, in alto) hanno risposto positivamente, affermando di riuscire a ricostruire il vincolo archivistico e proporre aggregazioni o relazioni tra i documenti, sulla base delle informazioni estratte dal contenuto dei documenti o dai metadati (compresi gli schemi di classificazione). Le risposte sono state di questo tipo: «our system contains clustering and ‘find similar’ technology which is able to propose aggregations and/or can be used to extract data that can be used for this purpose»; «yes, in most circumstances, but this is often contingent on the metadata available from the content source»; «any lost data can be regenerated if source-digital documents are available. The source digital documents can be used to extract the lost data»; «this re-construction can be performed [...], but limited to the cases where the aggregation logics can be extracted from the document text content»; «probably we can perform classification, and what this describes is ultimately a classification problem». Le altre otto aziende hanno risposto negativamente, aggiungendo che si tratta di un'operazione che può essere eseguita solo

⁵ Il vincolo o nesso archivistico è «l'insieme delle relazioni logiche e formali che esistono tra i documenti di un archivio» ([19]:68-69). Infatti, «il documento archivistico è prodotto nell'ambito di un'attività pratica, amministrativa o giuridica, e pertanto è strettamente correlato ai suoi precedenti e susseguenti, nonché al soggetto produttore [...]; esso è naturale e originario, nel senso che nasce spontaneamente senza una precisa volontà del soggetto produttore e si forma nel momento stesso in cui vengono prodotti i documenti» ([19]:69). Il vincolo archivistico «non è frutto di una volontà esterna, ma è naturale, caratterizzato dalla necessità e originarietà e che determina una lunga catena di precedenti susseguenti, catena che non può essere spezzata senza compromettere l'esistenza e l'essenza dell'archivio» ([7]:21).

manualmente, oppure che essi possono «provide restored records to users in the case of data loss at the content source, but cannot reconstitute aggregations».

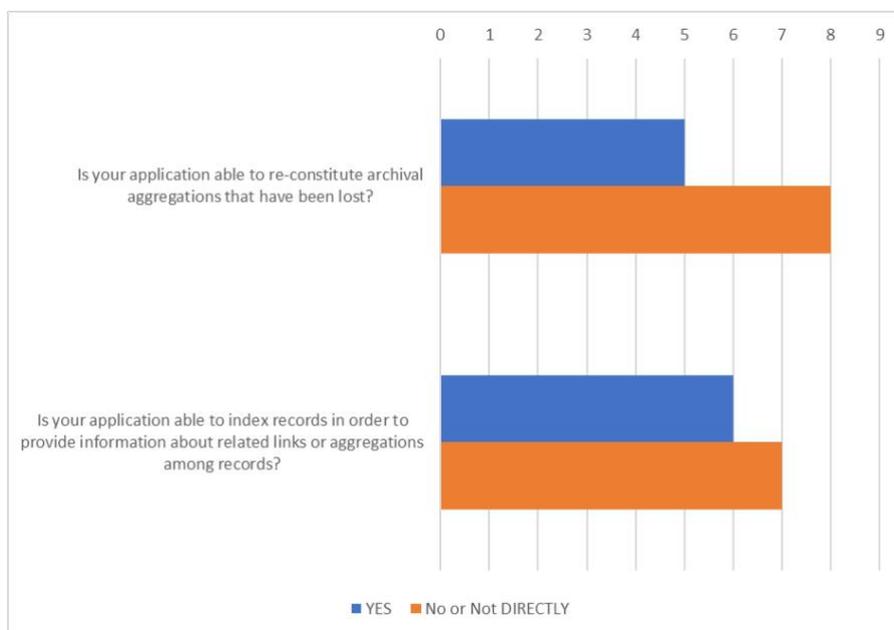


Figura 7. Domande e relative risposte sulla capacità di ricostituire il vincolo archivistico

Alle aziende intervistate è stato anche chiesto se le loro applicazioni fossero in grado di indicizzare i documenti per fornire informazioni sui collegamenti tra i documenti o sulle loro aggregazioni. Sei su tredici (vd. Figura 7, in basso) hanno risposto positivamente, con affermazioni del tipo: «we capture all contextual metadata including relationships between items and aggregations»; «all data in the system is always indexed»; «the technology does not natively work with record management, but can extract information from the text by understanding the content meaning and make it available to downstream record management systems which can index this information to create links, aggregations, ecc.». Le altre aziende hanno risposto in modo indeterminato o negativo, con risposte del tipo: «maybe, but maybe only in a very limited way»; «not explicitly, but if there is metadata available on the aggregation or record in the content source that will permit this then it is possible»; «this feature is in development at this time, but the application is able to identify similar emails based on indexing». Dalle risposte si può evincere come la ricostituzione delle aggregazioni documentali non sia l'obiettivo principale delle aziende che sviluppano tecnologie di intelligenza artificiale. Inoltre, si tratta di un compito molto difficile se si perdono le informazioni di contesto. Per questo sarà necessario attendere la disponibilità di tecnologie più evolute.

3.4 Estrazione e indicizzazione dei metadati

I metadati sono definiti nella letteratura scientifica semplicemente come “dati su (altri) dati”, «informazioni che descrivono un insieme di dati» ([12]:189). I metadati sono utilizzati per

molteplici scopi: ricerca, gestione e localizzazione, selezione, interoperabilità ([12]:189).⁶ Un indice è definito come un elenco di parole chiave associate a un documento, utilizzato in particolare come aiuto nella ricerca di informazioni. Lo scopo principale di un sistema di indicizzazione dei documenti «is to facilitate the indexing of records in the recordkeeping system, through assignment of access points to each record using a controlled recordkeeping vocabulary, for the purpose of facilitating effective and efficient discovery and retrieval of records in the recordkeeping system» ([11]:202).

Lo sviluppo di metodi e strumenti per l'acquisizione automatica di metadati normalizzati è fondamentale per l'integrità e l'autenticità delle fonti digitali. Le soluzioni di intelligenza artificiale possono aumentare l'arricchimento dei metadati e l'indicizzazione dei contenuti per espandere ulteriormente i collegamenti tra i concetti nelle ontologie e fornire un accesso ancora più semplice e veloce ai contenuti di interesse.

Tutte le aziende hanno sviluppato soluzioni per estrarre i metadati, ad esempio utilizzando *parser* di documenti o applicazioni OCR automatiche esterne per riempire i campi dei metadati del documento per descriverlo o per generare metadati dal contenuto, laddove l'OCR è possibile. Un'azienda ha risposto che «handwritten text can sometimes be effectively processed with software-based OCR tools, but not commonly» e che di solito «handwritten documents need to be OCRd by a hardware scanner which are much more sophisticated for this purpose». Alcune aziende hanno notato che la qualità del testo estratto attraverso l'elaborazione OCR non è elevata ed è costosa da gestire.

Esaminando più in dettaglio le risposte di alcune delle aziende intervistate, una di esse ha dichiarato che la sua applicazione è in grado di «extract information from scanned documents and this information may be used also to create records»; un'altra ha scritto che essa «can map metadata from most systems we connect to, can extract from the text of the document itself, and can perform ICR on handwritten text using a built-in Azure CV integration»; una terza azienda ha dichiarato che la sua piattaforma «scrapes extrinsic metadata from the content source and sometimes from intrinsic metadata from inside the record files» e che essa «also has the capability to add metadata defined by external systems using pattern matching rules. Some metadata is derived by AI technologies from the record text». Ha aggiunto, inoltre, che la sua applicazione «has a 'signaling' process in the enrichment pipeline that calculates new metadata based on the less refined metadata the system has harvested or mined»; una quarta azienda ha risposto che la sua applicazione «does not natively work on source handwritten documents, but can be integrated with existing tools that can perform the job (OCR, ICR, ecc.) thanks to a number of existing connectors, and that can work upstream. The output of these tools can be plugged to the technology components, so they can use the generated text, and possibly identify the document structure where needed to properly generate metadata, tags, labels, ecc. for the business purpose». Infine, due aziende hanno affermato che possono utilizzare le tecniche di Natural Language Understanding (NLU) per l'estrazione e l'indicizzazione dei metadati. L'NLU può consentire un'indicizzazione precisa e dettagliata e l'apposizione di tag semantici per rendere i contenuti ancora più accessibili e utilizzabili dagli esseri umani. Consentendo la creazione di un maggior numero di collegamenti tra argomenti e concetti, l'NLU permette di far emergere tutto

⁶ Come ha sottolineato Giovanni Michetti ([16]:269) «la gestione delle norme quantità di documenti prodotti in ambiente digitale richiede l'adozione di un robusto sistema di metadati a sostegno dei processi di creazione, trattamento e conservazione dei documenti. [...] I metadati documentano il contesto di produzione degli oggetti documentali e pertanto devono essere individuati in funzione del contesto stesso, oltre che delle specifiche finalità che ne giustificano la creazione e l'uso».

il valore delle informazioni, facilitando ai ricercatori il recupero di contenuti rilevanti e l'identificazione di nuove connessioni tra molteplici elementi, argomenti e metadati.

4. Considerazioni finali

Le risposte al questionario hanno messo in luce alcuni tratti comuni sui quali è interessante riflettere brevemente.

Innanzitutto, si evidenzia una comune consapevolezza del ruolo fondamentale dei metadati originali creati dal soggetto produttore durante le attività correnti, sia nel caso in cui l'obiettivo sia la classificazione automatica dei documenti, sia nel caso della creazione di aggregazioni archivistiche. Naturalmente, questa percezione è ancora più forte quando l'obiettivo è la ricostituzione di contesti archivistici originali. Per questo motivo, la presenza di eventuali metadati trovati o dedotti sui documenti è un elemento ricorrente in quasi tutte le risposte. Anche le informazioni sulla tipologia dei documenti – quando disponibili – sono considerate un altro componente cruciale per il successo dell'operazione.

Per quanto riguarda la classificazione dei documenti, solo un'azienda ha sottolineato la capacità della propria piattaforma di essere addestrata dagli utenti, grazie a uno specifico set di dati per generare autonomamente etichette e tag relativi a qualsiasi schema di classificazione dei documenti basato su tassonomie o ontologie. In tutti gli altri casi, l'intermediazione umana sembra ancora inevitabile se si vogliono fornire risultati coerenti.

Per quanto riguarda l'aggregazione dei documenti, le promesse di automazione non sono molto incoraggianti, in quanto questa possibilità è limitata a casi molto specifici. Le soluzioni disponibili sono in grado di aggregare in modalità automatica o semi-automatica i documenti basandosi sul loro contenuto ma ad oggi è sempre necessario un intervento di validazione umano. Inoltre, nella maggior parte dei casi queste soluzioni non sono del tutto mature ma ancora in fase di sviluppo.

Anche le informazioni sulla provenienza non sembrano essere facilmente identificabili dalle soluzioni di intelligenza artificiale quando si basano solo su meccanismi di inferenza e senza l'ausilio di informazioni di supporto (come l'identificazione del corretto fascicolo, la presenza di un timbro, una segnatura o di indicazioni chiaramente espresse nel documento, la disponibilità di metadati specifici e/o elementi di classificazione).

Allo stesso modo la ricostituzione del vincolo archivistico – elemento fondamentale delle aggregazioni documentali ([20]) – quando non è stato mai presente o è andato perduto, non è un'operazione semplice né realizzabile solamente affidandosi ad applicazioni di intelligenza artificiale; è sempre necessaria la disponibilità di ulteriori informazioni sulla struttura delle aggregazioni e/o un aiuto significativo da parte di operatori umani.

Un altro elemento importante che è stato messo in luce dallo studio è la 'sfida' della terminologia. Quando si interagisce con le aziende che sviluppano prodotti basati sull'intelligenza artificiale, si ha normalmente a che fare con professionisti informatici, che il più delle volte attribuiscono a determinati termini un significato diverso da quello attribuito loro dagli archivisti. Questo è il caso, ad esempio, di parole come 'classificazione', 'indicizzazione', 'clustering' che hanno un significato diverso nei due domini. Allo stesso modo, espressioni come 'Intelligent Document Processing', che si trovano sui siti di alcune aziende, non hanno nulla a che fare con il documento, il suo trattamento e, in ultima analisi, con l'intelligenza archivistica. Quindi, affinché

il confronto con le aziende risulti davvero proficuo, è bene dedicare del tempo all'inizio a chiarire i concetti di base e prestare molta attenzione affinché essi vengano interpretati correttamente.

In conclusione, quello che è emerso dallo studio è che allo stato attuale, nonostante i promettenti annunci pubblicitari presenti su molti dei siti web di aziende che propongono applicazioni di intelligenza artificiale per la gestione dei documenti, i risultati non sono ancora così lusinghieri. Tuttavia, in alcuni casi le aziende hanno dichiarato che queste funzionalità non sono ancora realmente efficaci perché ancora in fase di sviluppo, ma potrebbero diventarlo a breve. Ovviamente, non è possibile sapere se nel futuro si potranno effettivamente ottenere risultati migliori, ma quello che è certo è che questi saranno ottenuti solo al prezzo di maggiori investimenti, che, tuttavia, ad oggi non sembrano essere sostenuti da una corrispondente richiesta da parte del mercato.

5. Sviluppi futuri

Lo studio condotto testimonia che, almeno per il momento, la complessità delle funzioni archivistiche non può essere delegata ad un approccio completamente automatico basato sulle tecniche di intelligenza artificiale; tutt'al più può essere supportata da queste tecnologie, ma sempre sotto la supervisione o l'intermediazione di archivisti ed utenti.

Non è possibile affermare, senza ulteriori analisi e casi di studio, quale grado di intermediazione professionale sarà necessario in futuro, almeno in una prospettiva a breve o medio termine. Sono necessarie ulteriori indagini per valutare e misurare la qualità, l'affidabilità e l'accuratezza dei nuovi strumenti di intelligenza artificiale e per verificare quanto sia vera la promessa che viene da più parti annunciata come imminente, ovvero quella di poter svolgere automaticamente compiti quali la classificazione, l'aggregazione dei documenti, l'individuazione dei metadati, sostituendo del tutto gli operatori umani. Per questi motivi il gruppo di lavoro CU05 sta proseguendo l'attività di ricerca attraverso l'analisi di alcuni casi di studio proprio con l'obiettivo di testare in situazioni concrete le effettive possibilità offerte dalle soluzioni di intelligenza artificiale che evolvono costantemente e incrementano giorno dopo giorno le proprie capacità.

6. Ringraziamenti

Ringrazio i membri del gruppo di lavoro CU05 del progetto InterPARES TRUST AI, in particolare Maria Mata Caravaca, Massimiliano Grandi, Mariella Guercio e Bruna La Sorda che, insieme a me, hanno contribuito alla realizzazione dello studio.

Ringrazio anche Alan Pelz-Sharpe, Andrew Warland, James Lappin, Jenny Bunn e Paul Young per l'indispensabile aiuto fornito nell'individuazione del primo elenco di 300 aziende informatiche che sviluppano prodotti di intelligenza artificiale applicata al settore della gestione documentale.

Riferimenti

- [1] Agenzia per l'Italia Digitale. 2021. *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici*. (17 maggio 2021). https://www.agid.gov.it/sites/agid/files/2024-05/linee_guida_sul_documento_informatico.pdf.
- [2] Agenzia per l'Italia Digitale. 2021. *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici*. (17 maggio 2021). Allegato 5 (“Metadati”). https://www.agid.gov.it/sites/agid/files/2024-06/allegato_5_metadati.pdf.
- [3] Allegrezza, Stefano. 2022. *La conservazione degli archivi di posta elettronica*. Torre del Lago Puccini. Civita Editoriale.
- [4] Bonfiglio Dosio, Giorgetta. 2023. *Primi passi nel mondo degli archivi*. 5 ed. Padova. CLEUP.
- [5] Caravaca, Maria Mata. 2017. *Policies and Requirements for Archival Sedimentation in a Hybrid Records Management Environment: A Critical Analysis of International Writings*, PhD Thesis, Sapienza Università di Roma. <https://hdl.handle.net/11573/937637>.
- [6] Carucci, Paola, e Guercio, Maria. 2008. *Manuale di Archivistica*. Roma. Carocci. p. 209.
- [7] De Felice, Raffaele. 1998. *L'archivio contemporaneo. Titolario e classificazione sistematica di competenze nei moderni archivi di enti pubblici e privati*. Roma. La Nuova Italia Scientifica.
- [8] Decreto del Presidente della Repubblica 28 dicembre 2000, n. 445. *Testo unico delle disposizioni legislative e regolamentari in materia di documentazione amministrativa*. (GU n.42 del 20-02-2001 - Suppl. Ordinario n. 30).
- [9] Decreto Legislativo 7 marzo 2005, n. 82. *Codice dell'amministrazione digitale*. (GU n.112 del 16-05-2005 - Suppl. Ordinario n. 93).
- [10] DLM Forum Foundation. 2011. *MoReq2010: Modular Requirements for Records Systems*, Volume 1, “Core Services & Plug-in Modules”, Version 1.1. https://moreq.info/files/moreq2010_vol1_v1_1_en.pdf.
- [11] Eastwood, Terry, Hans Hofman and Randy Preston. 2008. “InterPARES 2 Project Book. Part Five: Modeling Digital Records Creation, Maintenance and Preservation: Modeling Cross-domain Task Force Report” in «Archivi», a. III-n.2 (luglio - dicembre 2008), numero speciale – Relazioni finali del gruppo internazionale di ricerca InterPARES 2. Disponibile anche all'indirizzo http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_5_modeling_task_force.pdf.
- [12] Guercio, Maria. 2010. *Archivistica informatica*. Roma. Carocci.
- [13] International Council on Archives (ICA). 2008. *ICA-Req: Principles and Functional Requirements for Records in Electronic Office Environments, Module 2: Guidelines and Functional Requirements for Electronic Records Management Systems*. pag.26 <https://www.ica.org/resource/ica-req-implementation-guidance-and-training-products>
- [14] InterPARES Trust AI project. <https://interparestrustai.org>.

- [15] InterPares Trust AI, CU05 working group. 2023. *The role of AI in identifying or reconstituting archival aggregations of digital records and enriching metadata schemas*, https://interparestrustai.org/assets/public/dissemination/Report-CU05-Survey-of-the-Companies_v121.pdf.
- [16] Michetti, Giovanni. 2014. “Gli standard di gestione documentale”, in (Giuva, Linda, e Maria Guercio, a cura di) *Archivistica. Teorie, metodi e pratiche*. Roma. Carocci.
- [17] Penzo Doria, Gianni. 2007. «Il fascicolo archivistico: le cinque tipologie e i modelli archivistici», in *Archivi e computer*, anno XVII, Fasc. 2-3/07.
- [18] Penzo Doria, Gianni. 2022. “Il piano per la fascicolatura”. In *Annuario dell'Archivio di Stato di Milano*. Milano. Archivio di Stato di Milano.
- [19] Pigliapoco, Stefano. 2022. *Documentare archiviare e conoscere. Formare e conservare la memoria nel contesto digitale*. Torre del Lago Puccini. Civita editoriale.
- [20] Romiti, Antonio. 2020. *Archivistica generale. Primi elementi*. Torre del Lago Puccini. Civita Editoriale.