

Archival users and AI tools for reference and access: a study within the InterPARES Trust AI project

Pierluigi Feliciati^(a)

a) University of Macerata, <https://orcid.org/0000-0002-2499-8528>

Contact: Pierluigi Feliciati, pierluigi.feliciati@unimc.it

Received: 25 April 2023; **Accepted:** 14 June 2023; **First Published:** 15 September 2023

ABSTRACT

Evaluating the effectiveness of search tools, the ease of use of the selection and display interfaces of archival descriptions and documents and the satisfaction of end users are not everyday activities among archive service managers. A culture of quality should instead consider actual effectiveness, adopting a user-centred design and a systematic comparison with final users. Within the international project InterPARES Trust AI, focused on the impact of Artificial Intelligence on archival functions, specific studies have been activated on user access to archives and documents in the AI perspective. This contribution presents this project, its assumptions, objectives, results and the first findings of the study in Italy.

KEYWORDS

User studies; Archives; Artificial Intelligence; InterPARES.

Gli utenti archivistici e gli strumenti IA per l'orientamento e l'accesso: uno studio nell'ambito del progetto InterPARES Trust AI

ABSTRACT

Valutare l'efficacia degli strumenti di ricerca, la facilità d'uso delle interfacce di selezione e visualizzazione delle descrizioni e dei documenti archivistici e la soddisfazione degli utenti finali non è un'attività diffusa tra i responsabili dei servizi archivistici. Una cultura della qualità dovrebbe invece tenere conto dell'efficacia reale, adottando una progettazione centrata sugli utenti e un confronto sistematico con essi. Nell'ambito del progetto internazionale InterPARES Trust AI, concentrato sull'impatto dell'Intelligenza Artificiale sulle funzioni archivistiche si sono attivati specifici studi sull'accesso degli utenti agli archivi e ai documenti, nella prospettiva dell'intelligenza artificiale. In questo contributo si presenta questo progetto, i suoi presupposti, obiettivi e risultati e i primi esiti dello studio in Italia sugli utenti.

PAROLE CHIAVE

Utenti; Archivi; Intelligenza Artificiale; IntePARES.

Premessa¹

Di cultura della qualità del servizio archivistico non si discute mai abbastanza, e non solo in Italia. Valutare l'efficacia degli strumenti di ricerca, la facilità d'uso delle interfacce di selezione e visualizzazione delle descrizioni e dei documenti archivistici e la soddisfazione degli utenti finali è un tema sullo sfondo del dibattito, come se non fosse prioritario. La corretta produzione, gestione e conservazione degli archivi sono presupposti imperativi della nostra professione, così come le loro condizioni di accesso rappresentano un obiettivo altrettanto centrale. Si documentano i fatti per poterne verificare, nel tempo, in modo autentico e accurato la sostanza e le implicazioni, conservandone gli atti, dopo accurata selezione e gestione, per prolungarne l'uso anche nei diversi, mutati contesti. Le funzioni archivistiche, nel loro diacronico articolarsi e nella dialettica con le attività della società, sono per ben noti motivi tutte cruciali e interdipendenti. Relegare la funzione di mediazione e accesso ai documenti dietro le quinte, riducendola al solo orario di apertura degli istituti di conservazione o alla mera pubblicazione in rete di inventari o banche dati, di documenti digitali o digitalizzati umilia il valore della descrizione, credo, contribuendo alla marginalità degli archivi nella percezione collettiva e dei *policy makers* (Felicati 2016).

Nella progettazione dei servizi archivistici digitali si declina spesso il requisito della qualità solo come rispetto ortodosso degli standard interni, con un approccio dunque auto-referenziale che resiste al confronto con l'efficacia percepita e con la soddisfazione. La mediazione archivistica va concepita come un processo aperto, capace di ripensarsi sia attraverso il confronto con le dinamiche (anche tecnologiche) della società, sia con le esigenze e i comportamenti degli "altri", degli utenti finali (Burton 2005; Featherstone 2006). L'attenzione alla qualità non può essere episodica perché le forme e i modi di gestione e uso degli archivi e dei documenti mutano al mutare degli strumenti disponibili per l'accesso alla conoscenza registrata (Daniels e Yakel 2010).

Muovendo da questi presupposti teorici ed etici è stato inevitabile attivare specifici studi sull'accesso degli utenti agli archivi e ai documenti nell'ambito del progetto internazionale InterPARES Trust AI, concentrato sull'impatto dell'Intelligenza Artificiale sulle funzioni archivistiche (Duranti et al. 2021). Di questo importante progetto e dei primi risultati dello studio in Italia condotto da chi scrive si tratta in questo contributo, sperando sarà di ispirazione l'approccio "collaborativo" con gli utenti, utile a fronteggiare senza pregiudizi ed esitazioni i cambiamenti che, ci piaccia o no, stiamo già vivendo.

Il progetto InterPARES Trust AI: storia, obiettivi e primi risultati

Nell'aprile 2021, il *Social Sciences and Humanities Research Council of Canada* ha assegnato una sovvenzione di partenariato al progetto quinquennale I Trust AI (dove I sta per InterPARES, *International Research on Permanent Authentic Records in Electronic Systems*, Trust evidenzia il concetto archivistico chiave dell'affidabilità e AI sta per Artificial Intelligence).² L'obiettivo generale è sviluppare e sfruttare soluzioni di IA per supportare la continua disponibilità e accessibilità di documenti pubblici autentici e affidabili. È stato costituito un consorzio internazionale che

¹ Le traduzioni dall'inglese nelle citazioni sono dell'autore.

² Ultimo accesso 24 aprile 2023. <https://interparestrustai.org/trust>.

copre quattro continenti per produrre ricerche originali e formare studenti e personale altamente qualificato, attivando un circolo virtuoso tra il mondo accademico, le istituzioni archivistiche, gli archivisti degli enti pubblici e l'industria, per rafforzare le conoscenze e le capacità di ciascuna delle parti. Coordinatrice del progetto è Luciana Duranti dell'Università della British Columbia, affiancata da Muhammad Abdul-Mageed e da Corinne Rogers.

Questo progetto rappresenta la quinta fase di InterPARES: la prima (1999-2001)³ era incentrata sul tema della conservazione dell'autenticità dei record elettronici non più necessari all'ente produttore per adempiere al proprio mandato; la seconda (2002-2007) ha modellizzato i concetti, i principi, i criteri e i metodi necessari per garantire la creazione e il mantenimento di documenti accurati e affidabili e la loro conservazione a lungo termine utilizzando le tecnologie informatiche, considerando le allora nuove tendenze all'interattività e alla partecipazione. La terza fase (2007-2012) ha tradotto la teoria e i metodi di conservazione digitale standard in piani d'azione concreti per supportare gli archivi di organizzazioni dotate di risorse limitate, evidenziando quali fattori influenzino l'implementazione appropriata per ciascun archivio in ciascun contesto e quali siano le competenze dei professionisti. Infine, InterPARES 4 (2012-2019)⁴ ha affrontato l'autenticità e affidabilità di record e dati in ambienti on-line, definendo i quadri teorici e metodologici utili per sviluppare politiche, procedure, regolamenti, standard e legislazioni locali, nazionali e internazionali.

Supportare il lavoro archivistico utilizzando l'IA non è un'idea nuova: da molto si discetta di *archivista cibernetico* (Rhoads 1969) e di *Artificial Intelligence, Expert Systems, and Archival Automation* (Hirtle 1987); Stielow paragonò lo sviluppo della teoria archivistica all'avanguardia dell'IA, caratterizzando entrambi come la ricerca dell'ovvio (Stielow 1991), mentre Benson adottò soluzioni di intelligenza artificiale per rappresentare le fotografie d'archivio (Benson 2009). Eppure, stando a una ricerca di Cohasset Associates e ARMA International, nel 2019 solo il 44% degli archivisti testimoniava che le proprie organizzazioni utilizzano “strumenti automatizzati [...] per individuare e conservare le informazioni rilevanti”, il 49% eliminava manualmente le e-mail, mentre il 58% eliminava manualmente i record dai dispositivi mobili (Cohasset e ARMA 2019). Le organizzazioni stanno annegando nei documenti e nelle informazioni, ormai è indiscutibile (Smallwood 2015). La situazione non è migliore quando le risorse documentarie sono innovative rispetto alle forme classiche di documenti archivistici. I messaggi di posta elettronica, ad esempio, possono essere fondamentali per documentare le responsabilità, l'applicazione dei diritti utili alla conservazione della memoria storica. “Tuttavia, conservare la posta elettronica è difficile e pone una serie di sfide logistiche ed etiche. Gli archivi hanno faticato a sviluppare strategie e metodi riproducibili sia per proteggere le informazioni sensibili nella posta elettronica, riservate o legalmente riservate, sia per garantire un accesso effettivo alla posta elettronica, a causa del volume e della complessità del materiale” (Schneider et al. 2019, 305-306). Allo stesso modo, i metadati – indispensabili per garantire nel tempo l'identità e l'integrità dei record – sono praticamente impossibili da gestire manualmente: “per gestire la quantità di metadati necessaria, gli archivisti e i gestori documentari dovranno quasi certamente usare sistemi di intelligenza artificiale che analizzino automaticamente la forma, il contenuto e il contesto degli oggetti che popolano i sistemi descrittivi” (Yeo 2013, 221).

³ Ultimo accesso 24 aprile 2023. <http://www.interpares.org/>.

⁴ Ultimo accesso 24 aprile 2023. <http://interparestrust.org/>.

Da quando Anne Gilliland ha esplorato per la prima volta l'utilizzo di un sistema esperto per classificare i record digitali (Gilliland 2016), si sono succeduti vari progetti che hanno esaminato l'IA per gli archivi; tuttavia, tali progetti sono in genere limitati, esaminando spesso un particolare strumento o anche un singolo set di record in un contesto specifico. Sebbene abbiano portata limitata, valutando solo una o due tecniche di intelligenza artificiale, gli studi generalmente concordano sul potenziale degli strumenti di intelligenza artificiale per la classificazione del contenuto di grandi raccolte di documenti e per supportare i gestori di record e gli archivisti nel loro lavoro quotidiano (Lee 2018). Oltre che per alleggerire il carico dei professionisti, alcuni studi hanno considerato l'adozione dell'intelligenza artificiale per arricchire l'accesso alle informazioni archivistiche, ad esempio con sistemi di raccomandazione che colleghino le immagini ai documenti digitalizzati (Vowell et al. 2020), utilizzando il riconoscimento del testo scritto a mano (*Handwritten Text Recognition* - HTR) per rendere ricercabili i documenti manoscritti (Goudarouli et al. 2019) e persino usando *chatbot* per aiutare i ricercatori a individuare collegamenti non espliciti (Gupta e Kapoor 2020).

La conoscenza sviluppata nei venti anni di ricerca di InterPARES costituisce un punto di partenza, pur con lo spostamento attuale dal focus sui progettisti di sistemi di record e archivi allo sviluppo dell'efficienza e dell'efficacia nello svolgimento delle funzioni archivistiche, proteggendo al tempo stesso la professione, l'impatto sociale e i valori culturali. L'approccio di ITrust AI è infatti al tempo stesso di breve e lungo termine: da un lato si pone la necessità pratica e urgente di affrontare i problemi esistenti su larga scala, identificandoli e valutando la possibile applicazione di soluzioni di IA per affrontarli coinvolgendo esperti e professionisti dell'archivistica e ricercatori di intelligenza artificiale su scala globale. Secondariamente, a lungo termine si esplorerà l'uso di strumenti di IA applicabili in modo affidabile per i problemi futuri, definendo gli strumenti di cui gli archivi avranno bisogno per soddisfare in modo flessibile le loro esigenze, in continua evoluzione, incluso il supporto per scegliere quali soluzioni basate sull'intelligenza artificiale saranno utili, opportune e sostenibili. Queste azioni devono considerare i contesti etico, legale e le implicazioni sociali di qualsiasi soluzione tecnologica per creare, gestire, utilizzare, controllare intellettualmente, conservare e fornire accesso agli archivi. Si devono valorizzare le promesse dell'IA evitando conseguenze indesiderate per l'integrità dell'infrastruttura dei documenti pubblici e applicando tali soluzioni per supportare la *governance* pubblica, la gestione, accesso e conservazione dell'evidenza documentaria e della cultura.

Parlando di intelligenza artificiale, durante la ricerca si stanno considerando una varietà di metodi di apprendimento automatico. Innanzitutto, i modelli di NLP (*Natural Language Processing*), che possono essere applicati al testo dei documenti e degli strumenti di gestione e accesso agli archivi: classificazione, traduzione, rilevamento di forme dialettali e persino identificazione di espressioni offensive o dannose (Zhang e Abdul-Mageed 2019; Elmadany et al. 2020). I metodi di *Deep Learning*, sia tradizionali che profondi, possono essere utilizzati per risolvere nuovi problemi archivistici: ad esempio, per migliorare gli strumenti di OCR (*Optical Character Recognition*) per garantire una migliore accessibilità e fruibilità dei documenti si aprono alcune questioni che vanno dall'analisi statistica degli *output* alla diversità linguistica nei dati di addestramento dei software (Smith e Cordell 2018). Il progetto sta affrontando anche le architetture di DCNN (*Deep Convolutional Neural Networks*) per una classificazione delle immagini con la combinazione di diversi metodi. Sebbene le ricerche passate abbiano già esaminato alcune tecniche tradizionali di apprendimento automatico (le cosiddette *bag of words*), non sono state valutate finora le tecniche più

recenti di *Deep Learning*, con tecniche di apprendimento profondo come le DCNN (Patel et al. 2020), che mostrano risultati promettenti. Ad esempio, uno strumento utile potrebbe essere la descrizione automatica dei documenti archivistici: per le riproduzioni digitali di documenti prive di descrizioni, potrebbero essere utilizzati i modelli Faster R-CNN basati sull'attenzione (Anderson et al. 2018). Allo stesso modo, l'apprendimento *multitasking* può essere utilizzato per migliorare la qualità delle traduzioni da una lingua all'altra, il rilevamento di oggetti e la classificazione delle immagini (Elliott e Kádár 2017), associate alle tecniche di NLP per la generazione di nuovo testo (Fedus, Goodfellow, e Dai 2018), la sua classificazione (Cireşan, Meier, e Schmidhuber 2012) e traduzione automatica (Przystupa e Abdul-Mageed 2019).

Il progetto I Trust AI è articolato in cinque fasi:

1. (2021-22) identificare quali specifiche tecnologie di IA sono in grado di affrontare le sfide critiche per record e archivi;
2. (2022-23) determinare i rischi e i benefici dell'utilizzo di tali tecnologie;
3. (2023-24) stabilire in che modo i concetti e i principi archivistici possono supportare lo sviluppo di un'IA responsabile;
4. (2024-25) convalidare i risultati raggiunti nella terza fase attraverso studi di casi e dimostrazioni;
5. (2025-26) verificare e convalidare tutti i prodotti e completare la strategia di disseminazione.⁵

Il coordinamento generale è in capo a un *Executive Research Committee* che si riunisce ogni 15 giorni, valuta gli studi proposti⁶ e organizza ogni quattro mesi i *plenary meeting* di tutti i partner, cui segue una giornata di convegno pubblica, sempre in sedi diverse in tutto il mondo. Inoltre, sono stati istituiti sei gruppi di lavoro fondati sulle funzioni archivistiche, cui afferiscono gli studi specifici: *Creation and Use*, *Appraisal and Acquisition*, *Arrangement and Description*, *Retention and Preservation*, *Management and Administration* e *Reference and Access*. Per facilitare la trasversalità di alcuni studi e lo scambio dei risultati sono stati poi creati nel corso delle attività alcuni *clusters* tematici: *Digitization*, *Analysis-Sensemaking*, *Privacy and Personal Information*, *User focus*. Uno strumento importante del progetto, vista la sua dimensione globale, è la ricca banca dati terminologica, accessibile pubblicamente.⁷

Lo studio sugli utenti

Nell'ambito del gruppo di lavoro su *Reference and Access*, fin dall'autunno 2021 è stato approvato ed è partito uno studio internazionale sugli utenti, identificato dalla sigla RA05 e dal titolo *Users' approaches and behaviors in accessing records and archives in the perspective of AI: a global user*

⁵ Ultimo accesso 24 aprile 2023. https://interparestrustai.org/trust/about_research/summary.

⁶ Tutti gli studi del progetto sono consultabili online. Ultimo accesso 24 aprile 2023. https://interparestrustai.org/trust/about_research/studies.

⁷ Ultimo accesso 15 aprile 2023. <https://interparestrustai.org/terminology>. La sezione del progetto si compone di un *Glossario*, di un Dizionario, di una sezione con le ontologie prodotte nei vent'anni di InterPARES e del rinvio al *Multilingual Archival Terminology* database (MAT) dell'International Council of Archives, coordinato da Luciana Duranti. Ultimo accesso 15 aprile 2023. <http://www.ciscra.org/mat/>.

study, traducibile in italiano con “Approcci e comportamenti nell’accesso a documenti e archivi nella prospettiva dell’Intelligenza artificiale”. Mancano studi di livello internazionale sull’effettiva qualità nell’esperienza di accesso e uso degli archivi e conseguentemente non sono disponibili protocolli e metriche condivise. I pochi studi effettuati sono sempre di livello locale, limitati a specifici servizi gestiti da specifiche organizzazioni (Chapman 2010; Daniels e Yakel 2010; Duff e Stoyanova 1998; Feliciati 2012; Scheir 2005; Yakel 2004).

Lo studio mira ad acquisire dati attendibili sul punto di vista degli utenti, così da poter supportare in ITrust AI la definizione di requisiti e di linee guida per lo sviluppo di strumenti di intelligenza artificiale affidabili, sostenibili e utili per migliorare le funzioni di supporto e accesso a documenti e archivi. Le questioni di partenza sono: cosa sappiamo di come gli utenti eseguono le proprie ricerche? Usano nomi personali, denominazioni di soggetti produttori e di uffici specifici, di luoghi? Sono a loro agio con il linguaggio delle interfacce digitali e dei documenti? Sono disposti a far elaborare i loro dati di ricerca da strumenti di IA per migliorare i servizi di ricerca e accesso? Apprezzeranno interfacce basate sul linguaggio naturale, anche tramite dispositivi audio? Sono individuabili tendenze e problematiche simili in diversi contesti d’uso, a livello internazionale?

I dati raccolti attraverso le attività di coinvolgimento di campioni di utenti finali riveleranno preziose informazioni sulla soddisfazione nei confronti dei servizi di riferimento e accesso agli archivi digitali esistenti e sulla effettiva consapevolezza, sulle aspettative e le preoccupazioni rispetto all’applicazione dell’IA. Il progetto è coordinato da chi scrive e condotto con Giorgia Di Marcantonio, Emanuele Frontoni e Marina Paolanti (UniMC), Jessica Bushey (USA), Thsepho Mosweu (Botswana), Darra Hoffmann (USA), Adele Torrance (Canada), Sindiso Bhebhe (South Africa).

La prima fase dello studio in Italia: metodologia e risultati

Prima di tutto nell’ambito dello studio RA05 è stata raccolta una bibliografia tematica sugli studi degli utenti archivistici, per avere sotto controllo lo stato delle ricerche e delle metodologie, coordinata con la bibliografia condivisa da tutti i ricercatori e i partner di ITrust AI, alimentata in Zotero⁸ per supportare gli studi in corso.

Al tempo stesso si è elaborato un protocollo di indagine per l’organizzazione di focus group, in inglese, condiviso con il gruppo di lavoro, che lo ha perfezionato nella prospettiva di adottare la stessa metodologia di raccolta dati nei diversi contesti geografici e archivistici. Si è dovuto però constatare che l’organizzazione di focus group sarebbe stata troppo impegnativa per i ricercatori e le istituzioni archivistiche coinvolte. Si tratta, a parere di chi scrive, di un segnale preoccupante del distacco esistente tra gli istituti archivistici e i loro utenti finali, in parte forse imputabile alla abitudine all’interazione a distanza dovuta alla difficile fase della pandemia da covid-19, forse alla rarità di interazioni mirate alla valutazione della qualità dei servizi, reali o on-line che siano. Pertanto, si è proceduto a rielaborare il protocollo in una forma adatta alla progettazione di questionari online, sia in inglese che in italiano, rivedendo l’agenda dell’intero studio convertendola dal percorso globale-locale – attivando diverse attività di studio a livello internazionale basate

⁸ Cfr. <https://www.zotero.org/>.

sullo stesso protocollo per poi raccogliere i dati e intervenire successivamente a personalizzazioni di contesto – all’opposto, locale-globale, testando la validità del protocollo nell’ambito di uno specifico studio (in Italia, in particolare), assestandone le criticità e applicandolo in un secondo momento su scala internazionale.⁹

La *survey* italiana basta sul protocollo condiviso con il team internazionale è stata lanciata il 10 ottobre 2022 grazie alla disponibilità dei direttori di quattro archivi di Stato italiani partner di I Trust AI, Ascoli Piceno, Milano, Torino e Venezia a provare a coinvolgere i propri utenti nella valutazione dei servizi digitali di *reference* e accesso nella prospettiva dell’adozione futura di strumenti di IA. Più in particolare, quando possibile, ci si è concentrati sull’accesso ai fondi relativi alle pratiche riguardanti le autorizzazioni all’uso dei cementi armati, a parere di chi scrive un caso esemplare di documentazione storica con un forte impatto sulla società contemporanea. Con il regio decreto 2229 del 1939 erano state definite le “Norme per l’esecuzione delle opere in conglomerato cementizio, semplice od armato”. In particolare, nell’art. 4, si fissava l’obbligo, per i privati, di deposito presso la Prefettura di una domanda corredata da un progetto di massima. Al termine dei lavori era obbligatorio un collaudo delle strutture da parte di un ingegnere incaricato, il cui certificato doveva essere presentato alla Prefettura stessa. Dagli anni ’70 del XX secolo questa funzione è stata delegata alle Regioni. Per l’intera provincia di Ascoli Piceno, ad esempio, si sono sedimentati 10.543 fascicoli (ora nel fondo Prefettura di Ascoli Piceno, *Cementi*, 1961 - 1975). La documentazione è stata versata in Archivio di Stato nel 2004 e dal 2016 è stato avviato - utilizzando un software dedicato - il lavoro di schedatura sistematica dei fascicoli, molto richiesti dai professionisti coinvolti nei cantieri di ricostruzione seguiti allo sciame sismico del 2016.¹⁰ Per Milano, oltre all’accesso alla serie della Prefettura, *Ufficio controllo opere in cemento armato*, 35.677 fascicoli (1947-1990), si è allargata la valutazione degli utenti della loro esperienza con la sezione web delle “riproduzioni digitali in rete e Banche dati”, riferita a circa 50 fondi.¹¹ Per l’archivio di Stato di Torino è stata proposta la “Sala di studio virtuale”, che riguarda la maggior parte del patrimonio conservato, inclusa la serie della Prefettura, *Pratiche cemento armato*, 48.000 fascicoli (1955-1971).¹² Infine, per Venezia, si è chiesto agli utenti di commentare la propria esperienza d’uso con *Moreveneto*, il sistema informativo dell’Archivio di Stato.¹³

Il questionario, mantenuto attivo fino all’11 dicembre, ha raccolto in totale 33 risposte, il maggior numero delle quali dagli utenti degli Archivi di Milano (45,5%) e di Ascoli Piceno (45,5%). Una quota minore di risposte è pervenuta da Venezia (circa il 6%) e da Torino (4%). La prima sezione del questionario mirava a raccogliere informazioni sugli utenti, la maggior parte dei quali sono risultati essere archivisti o bibliotecari (21,2%), insegnanti (21,2%), oppure professionisti di vari settori, come l’architettura e l’ingegneria (33,3%). I partecipanti al sondaggio sono stati prevalentemente (oltre il 65%) di età compresa tra 35-44 anni oppure oltre i 65 anni.

La seconda sezione del questionario era pensata per indagare come gli utenti effettuano ricerche di informazioni online. Le risposte hanno rivelato che molti di loro (26%) si affida spesso o sempre

⁹ Per tutte queste attività è stata indispensabile la collaborazione di Giorgia Di Marcantonio, che ringrazio.

¹⁰ Ultimo accesso 15 aprile 2023. <https://www.archiviodistatoap.it/cementi-servizi-online>.

¹¹ Ultimo accesso 15 aprile 2023. <https://archiviodistatomilano.cultura.gov.it/patrimonio/banche-dati>.

¹² Ultimo accesso 15 aprile 2023. <https://archiviodistatorino.beniculturali.it/sala-studio-virtuale/>.

¹³ Ultimo accesso 15 aprile 2023. <https://asve.arianna4.cloud/>.

ai motori di ricerca generalisti, con solo una minima percentuale (1%) che li utilizza sfruttandone le funzioni avanzate. Inoltre, oltre l'87% degli intervistati ha riferito di utilizzare per le proprie *query* parole o frasi come chiavi di ricerca (ad es. "Archivi dell'ufficio finanziario"), mentre oltre l'81% fa uso della ricerca di immagini di Google. Questa tendenza all'uso di risorse non specializzate si estende anche alle ricerche di informazioni archivistiche, con la maggior parte degli utenti che optano per i motori di ricerca generalisti piuttosto che per i sistemi archivistici nazionali. In caso, gli utenti più specializzati (circa il 20%) preferiscono ricorrere direttamente ai siti degli Archivi di Stato.

In merito all'esperienza degli specifici servizi archivistici proposti, i partecipanti hanno espresso il desiderio di una maggiore disponibilità di risorse digitalizzate (35,5%) e di un miglioramento complessivo della qualità delle interfacce nel reperimento delle informazioni (29%). Inoltre, alcuni (12,9%) auspicano una maggiore facilità di accesso alle risorse fisiche conservate nell'archivio. Purtroppo, è risultato che non è spesso possibile per gli utenti prenotare online ciò che desiderano consultare fisicamente nelle sale di studio degli archivi e comunque la consultazione risulta limitata a poche unità. Nonostante queste limitazioni, oltre il 90% degli intervistati ha indicato che continuerebbe a utilizzare il servizio digitale proposto, indicando per esso un alto livello di soddisfazione.

La sezione finale del questionario era dedicata a indagare il rapporto tra gli utenti e possibili strumenti di intelligenza artificiale nei sistemi di ricerca archivistica. È interessante notare che la maggior parte degli utenti intervistati (91%) conosceva già l'esistenza dell'intelligenza artificiale, anche se 1 su 5 non è consapevole che le proprie attività in rete sono già processate automaticamente e solo il 63% ha espresso il proprio consenso ad incorporare l'intelligenza artificiale nei meccanismi di ricerca. Tuttavia, un quinto circa degli intervistati ha riconosciuto di non avere piena consapevolezza su cosa siano e cosa comportino gli strumenti di intelligenza artificiale quindi, in caso glielo si chiedesse, pretenderebbero maggiori informazioni sui vantaggi e gli svantaggi del loro utilizzo. Consenso particolare hanno raccolto l'adozione di strumenti di elaborazione automatica del testo dei documenti digitalizzati (75%) e l'uso del linguaggio naturale per l'interazione con gli strumenti di mediazione e accesso (68%).

Conclusioni

I risultati di questo studio forniscono informazioni preziose sulle esperienze e le aspettative degli utenti nell'accesso ai servizi archivistici digitali e sul loro atteggiamento nei confronti dell'adozione futura di soluzioni basate sull'intelligenza artificiale. L'indagine ha rivelato che il campione di intervistati italiani – faticosamente raggiunto, va sottolineato ancora una volta - tende ad affidarsi ai motori di ricerca generalisti per le ricerche online, pur non utilizzandone le funzioni avanzate. Sperano in una maggiore disponibilità di risorse digitalizzate, di servizi di prenotazione della consultazione in sala di studio e lamentano la qualità complessiva delle interfacce per il reperimento delle informazioni.

Inoltre, una parte consistente degli utenti ha espresso il proprio consenso sull'incorporare l'intelligenza artificiale negli strumenti di ricerca online, ma ha anche indicato al tempo stesso la necessità di avere informazioni puntuali sui possibili vantaggi e svantaggi. Questo risultato potrebbe

essere letto anche alla luce di una recente ricerca: “Il 78% degli intervistati cinesi (la percentuale più alta dei paesi intervistati) concorda con l’affermazione secondo cui i prodotti e i servizi che utilizzano l’IA offrono più vantaggi che svantaggi. Dopo gli intervistati cinesi, quelli dell’Arabia Saudita (76%) e dell’India (71%) si sono espressi più positivamente nei confronti dei prodotti di IA. Solo il 35% degli americani inclusi nel campione (tra i più bassi dei paesi intervistati) concorda sul fatto che i prodotti e i servizi che utilizzano l’intelligenza artificiale rivestano più vantaggi che svantaggi” (Stanford 2023, 4). L’Italia, in questo studio, ha raccolto il 50% di intervistati ottimisti, una percentuale inferiore rispetto al consenso del campione di utenti archivistici coinvolti nella ricerca che qui si presenta.

Nell’ambito del progetto ITrust AI, lo studio evidenzia l’utilità di ulteriori ricerche per comprendere meglio i comportamenti e la soddisfazione degli utenti nei confronti dei servizi di ricerca e di accesso agli archivi digitali, nonché una percezione incerta in merito ai potenziali vantaggi e svantaggi nell’adozione di strumenti di intelligenza artificiale per migliorarli. I risultati, che aumenteranno il proprio significato quando saranno confrontati e riuniti con i dati provenienti dalle ricerche svolte in altri paesi, potranno contribuire allo sviluppo di scelte più ponderate su quali strumenti di intelligenza artificiale siano funzionali e affidabili per rafforzare i servizi archivistici.

Riferimenti bibliografici

- Anderson, Peter, He Xiadong, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, e Lei Zhang. 2018. “Bottom-up and top-down attention for image captioning and visual question answering.” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6077-6086. <https://doi.org/10.1109/CVPR.2018.00636>.
- Belovari, Susanne. 2017. “Expedited digital appraisal for regular archivists: an MPLP-type approach.” *Journal of Archival Organization* 14(1-2): 55-77. <https://doi.org/10.1080/15332748.2018.1503014>.
- Benson, Allen C. 2009. “The Archival Photograph and Its Meaning: Formalisms for Modeling Images.” *Journal of Archival Organization* 7(4): 148–187. <https://doi.org/10.1080/15332740903554770>.
- Burton, Antoinette. 2005. *Archive Stories: Facts, Fictions, and the Writing of History*. Durham: Duke University Press. <https://doi.org/10.2307/j.ctv11smn7b>.
- Featherstone, Mike. 2006. “Archive.” *Theory Culture Society* 23(2-3): 591-596. <https://doi.org/10.1177/0263276406023002106>.
- Chapman, Joyce. 2010. “Observing Users: an Empirical Analysis of User Interaction with Online Finding Aids.” *Journal of Archival Organization* 8(1): 4-30. <https://doi.org/10.1080/15332748.2010.484361>.
- Cireşan, Dan C., Ueli Meier, e Jurgen Schmidhuber. 2012. “Transfer learning for Latin and Chinese characters with deep neural networks.” In *The Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN)* 1-6. <https://doi.org/10.1109/IJCNN.2012.6252544>.
- Cohasset Associates e ARMA International. (n.d.). 2019 *Information Governance Benchmarking Report*. Ultimo accesso 24 aprile 2023. https://armai.informz.net/ARMAI/pages/Cohasset_Benchmarking_Survey_2019.
- Daniels, Morgan, e Elizabeth Yakel. 2010. “Seek and You May Find: Successful Search in Online Finding Aid Systems.” *American Archivist* 73(2): 535-568. <https://doi.org/10.17723/aarc.73.2.p578900680650357>.
- Duff, Wendy, e Penka Stoyanova. 1998. “Transforming the Crazy Quilt: Archival Displays from User’s Point of View.” *Archivaria* 45: 44-79. https://www.researchgate.net/publication/261510221_Transforming_the_Crazy_Quilt_Archival_Displays_from_a_Users'_Point_of_View.
- Dunley, Richard. 2018. “Machines reading the archive: handwritten text recognition software.” *The National Archives Blogpost* (blog). Ultimo accesso 24 aprile 2023. <https://blog.nationalarchives.gov.uk/machines-reading-the-archivehandwritten-text-recognition-software/>.
- Duranti, Luciana, Muhammad Abdul-Mageed, Darra Hofman, e Peter Sullivan. 2021. “I Trust AI, the New InterPARES Research Project.” *Anuario Escuela de Archivología* 13: 36-55. <https://revistas.unc.edu.ar/index.php/anuario/issue/view/2522>.
- Elliott, Desmond, e Kádár, Akos. 2017. “Imagination improves multimodal translation.” *arXiv preprint*. <https://doi.org/10.48550/arXiv.1705.04350>.

- Elmadany, AbdelRahim, Chiyu Zhang, Muhammad Abdul-Mageed, e Azadeh Hashemi. 2020. “Leveraging Affective Bidirectional Transformers for Offensive Language Detection.” In *Proceedings of the Fourth Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT4)*. <https://doi.org/10.48550/arXiv.2006.01266>.
- Fedus, William, Ian Goodfellow, e Andrew M. Dai. 2018. “MaskGAN: Better text generation via filling in the ___.” *arXiv preprint*. <https://doi.org/10.48550/arXiv.1801.07736>.
- Feliciati, Pierluigi. 2012. “Ask the users, il valore aggiunto della valutazione dei sistemi informativi culturali on line coinvolgendo gli utenti: Il caso del progetto Una Città per gli Archivi.” *Il capitale culturale* 5: 129-144. <http://dx.doi.org/10.13138/2039-2362/483>.
- Feliciati, Pierluigi. 2016. “L’usabilità degli ambienti bibliotecari e archivistici digitali come requisito di qualità: contesto, modelli e strumenti di valutazione.” *JLIS.it* 7(1): 113-130. <http://dx.doi.org/10.4403/jlis.it-11512>.
- Gilliland, Anne J. 2016. “Designing Expert Systems for Archival Evaluation and Processing of Computer Mediated Communications: Frameworks and Methods.” In *Research in the Archival Multiverse*, a cura di Anne J. Gilliland, Sue McKemmish, AJ Lau. UCLA. <https://escholarship.org/uc/item/6c21p3hs>.
- Goudarouli, Eirini, Anna Sexton, e 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. <https://doi.org/10.1007/s13347-018-0333-3>.
- Gupta, Abhishek, e Nikitasha Kapoor. 2020. “Comprehensiveness of Archives: A Modern AI-enabled Approach to Build Comprehensive Shared Cultural Heritage.” <https://doi.org/10.48550/arXiv.2008.04541>.
- Hirtle, Peter B. 1987. “Artificial Intelligence, Expert Systems, and Archival Automation.” *Provenance, Journal of the Society of Georgia Archivists* 5(1): 76-88. <https://digitalcommons.kennesaw.edu/provenance/vol5/iss1/8>.
- Lee, Christopher A. 2018. “Computer-Assisted Appraisal and Selection of Archival Materials.” In *2018 IEEE International Conference on Big Data*, 2721–2724. <https://doi.org/10.1109/BigData.2018.8622267>.
- Makhlouf Shabou, Basma, Julien Tièche, Julien Knafou, e Arnaud Gaudinat. 2020. “Algorithmic methods to explore the utomation of the appraisal of structured and unstructured digital data.” *Records Management Journal* 30(2): 175–200. <http://dx.doi.org/10.1108/RMJ-09-2019-0049>.
- Przystupa, Michael, e Muhammad Abdul-Mageed. 2019. “Neural machine translation of low-resource and similar languages with backtranslation.” In *Proceedings of the Fourth Conference on Machine Translation* 3: 224-235. <http://dx.doi.org/10.18653/v1/W19-5431>.
- Rhoads, James B. 1969. “The Historian and the New Technology.” *The American Archivist* 32(3): 209–213. <https://doi.org/10.17723/aarc.32.3.f8181nm6mjv147t7>.
- Rolan, Gregory, Glenn Humphries, Lisa Jeffrey, Evanthisa Samaras, Tatiana Antsouпова, e Katharine Stuart. 2019. “More human than human? Artificial intelligence in the archive.” *Archives and Manuscripts* 47(2): 179–203. <https://doi.org/10.1080/01576895.2018.1502088>.

Schneider, Josh, Chance Adams, Sally DeBauche, Reid Echols, Callum McKean, Jessica Moran, e Dorothy Waugh. 2019. "Appraising, processing, and providing access to email in contemporary literary archives." *Archives and Manuscripts* 47(3): 305–326. <https://doi.org/10.1080/01576895.2019.1622138>.

Smallwood, Richard. 2015. *Information Governance: Concepts Strategies and Best Practices*. New York: Wiley.

Smith, David, e Ryan Cordell. 2018. *A research agenda for historical and multilingual optical character recognition*. NULab, Northeastern University Report. Ultimo accesso 24 aprile 2023. <https://ocr.northeastern.edu/report>.

Stanford HAI - Institute for Human-Centered Artificial Intelligence. 2023. *AI Index Report, 8, Public Opinion*. Ultimo accesso 24 aprile 2023. https://test-hai-aiindex.pantheonsite.io/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_8.pdf.

Stanford HAI - Institute for Human-Centered Artificial Intelligence. 2023. *Chapter 8: Public Opinion*. Artificial Intelligence Index Report 2023. Ultimo accesso 24 aprile 2023. https://test-hai-aiindex.pantheonsite.io/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_8.pdf.

Stielow, Frederick J. 1991. "Archival Theory Redux and Redeemed: Definition and Context Toward a General Theory." *American Archivist* 54: 14–26. <https://www.jstor.org/stable/40294399>.

Van Hooland, Seth, e Mathias Coeckelbergs. 2018. "Unsupervised Machine Learning for Archival Collections: Possibilities and Limits of Topic Modeling and Word Embedding." *Revista catalana d'arxivística* 41: 73-90. <https://arxiv.org/abs/2018.10.14> -Dossier [SVHooland_MCoeckelbergs.pdf](https://arxiv.org/abs/2018.10.14).

Vellino, André, e Inge Alberts. 2016. "Assisting the appraisal of e-mail records with automatic classification." *Records Management Journal* 26(3): 293-313. <https://doi.org/10.1108/RMJ-02-2016-0006>.

Vowell, Zach, Ethan Kusters, Luca Soares, Samuel Verkruyse, Joey Wilson, e Foaad Khosmood. 2020. *Morgan Papers: Exploring the Correspondence of California's First Female Architect*. Ultimo accesso 24 aprile 2023. <https://dblp.org/rec/conf/dihu/VowellKSVWK20.html>.

Yakel, Elizabeth. 2004. "Encoded Archival Description: Are Finding Aids Boundary Spanners or Barriers for Users?." *Journal of Archival Organization* 2(1-2): 63-77. <https://doi.org/10.1300/J201v02n0106>.

Yeo, Geoffrey. 2013. "Trust and context in cyberspace." *Archives and Records* 34(2): 214–234. <https://doi.org/10.1080/23257962.2013.825207>.

Zhang, Chiyu, e Muhammad Abdul-Mageed. 2019. "No army, no navy: Bert semi-supervised learning of Arabic dialects." In *Proceedings of the Fourth Arabic Natural Language Processing Workshop (OSACT4)*: 279-284. Ultimo accesso 24 aprile 2023. <https://aclanthology.org/W19-4637.pdf>.