

Titolo del progetto proposto:

L'Intelligenza Artificiale al servizio della cultura: digitalizzazione e recupero del Patrimonio Leopardiano

Descrizione sintetica della proposta di progetto (max 500 caratteri spazi esclusi)

Il progetto intende contribuire allo studio e all'implementazione di strategie di digitalizzazione e metadattazione di beni librari e archivistici conservati presso il Centro Nazionale di Studi Leopardiani, verificando altresì in che misura l'Intelligenza Artificiale (d'ora in avanti AI) possa coadiuvare l'attività di valorizzazione dei metadati amministrativi e gestionali per i beni digitalizzati, nonché favorire il recupero e il ripristino totale o parziale di file digitali degradati o compromessi a motivo dello stato di conservazione degli originali.

Identificazione del problema (max 1500 caratteri spazi esclusi)

I file digitali, pur se archiviati in formati idonei alla conservazione a lungo termine, sono soggetti a due principali criticità: la degradazione progressiva, dovuta alla naturale obsolescenza dei supporti digitali e delle strutture dati, e la compromissione nativa, legata al deterioramento o alla scarsa qualità dei documenti analogici da cui essi derivano. Questi fenomeni pregiudicano l'affidabilità dei file, rendendo necessarie continue e innovative strategie di recupero e preservazione dei dati. Parallelamente, la valorizzazione dei metadati METS – che costituisce un passaggio cruciale per garantire la tracciabilità, l'interoperabilità e la qualità delle collezioni digitali – richiede elevata precisione, risorse significative e operazioni ripetitive, rendendo auspicabile un'automatizzazione dei processi. Si intende dunque verificare in che modo l'AI possa essere implementata virtuosamente nelle attività di digitalizzazione e metadattazione di beni culturali, individuando come caso di studio la collezione bibliografica e archivistica del Centro Nazionale di Studi Leopardiani (d'ora in avanti CNSL). Essa, infatti, è composta da documenti cartacei e multimediali (audiocassette, videocassette, diapositive, ecc.) che il CNSL sta provvedendo a censire digitalmente in portali ministeriali (MOL, OPAC SBN, SIUSA/SIA) e che vorrebbe anche digitalizzare per favorirne diffusione e valorizzazione (con particolare urgenza nel caso dell'Archivio multimediale perché composto da dispositivi ormai obsoleti e attualmente preclusi alla consultazione a causa della mancanza presso il CNSL di attrezzature in grado di processarli).

Obiettivi generali e specifici del progetto e loro pertinenza con l'iniziativa Digital MAB

Gli obiettivi generali del progetto, intesi come traguardi di interesse diffuso per il contesto MAB, sono: 1) favorire lo studio di buone pratiche per la digitalizzazione dei beni culturali; 2) contribuire alla definizione di un modello di metadati amministrativi e gestionali specifico per i beni documentali e multimediali che possa essere condiviso e replicabile sul piano nazionale; 3) studiare le possibili forme di integrazione dell'AI al mondo dei beni culturali, applicandola alla valorizzazione automatizzata dei metadati e alle strategie di conservazione a lungo termine, anche tramite recupero di file digitali compromessi (o per alterazione dell'originale o per degradazione fisiologica dei file). Gli obiettivi specifici del progetto, intesi come traguardi di interesse distintivo per gli enti partner, sono: 1) digitalizzazione dei manoscritti posseduti dal CNSL e di una selezione dei volumi conservati in Biblioteca; 2) inventariazione, digitalizzazione e salvaguardia dell'Archivio multimediale del CNSL; 3) avvio di un processo di formazione specifica sui temi della convergenza digitale per maturare nuove conoscenze sul mondo digitale, affrontare con consapevolezza e competenza crescenti le sfide che esso propone, e sfruttare le sue opportunità di inclusione e democratizzazione dell'accesso alla cultura (con ricadute sulla comunità, ad esempio mediante strategie di crowdsourcing); 4) mettere a disposizione del pubblico, attraverso i portali ministeriali e la piattaforma Biblioteca Digitale Leopardiana, le collezioni del CNSL, facendole dialogare con tutte le altre risorse presenti (autografi, testi, ecc.) per favorire lo studio e la ricerca.

Declinazione del modello di ricerca-azione (illustrare le attività e le metodologie proposte per l'identificazione e l'implementazione di azioni e soluzioni alle problematiche rilevate)

Il modello di ricerca-azione pensato per il progetto prevede, da un lato, di attuare una bipartizione tematico-organizzativa che permetta di considerare separatamente le due aree di intervento (cartacea e multimediale) per rispettare le diverse specificità dei beni in sede di digitalizzazione e metadattazione; dall'altro, di favorire un ricongiungimento tecnico-procedurale delle due aree per lo studio del tracciato di metadati da associare ai file digitali e il loro conferimento alla Digital Library nazionale. Per la prima area di intervento (cartacea) si prevedono le seguenti attività: 1) studio dello schema di metadati METS e identificazione di un modello descrittivo uniforme per i file digitali; 2) digitalizzazione dei supporti attraverso la strumentazione già in possesso del CNSL secondo gli standard del Piano Nazionale per la Digitalizzazione del Patrimonio Culturale (d'ora in avanti PND). Per la seconda area di intervento (multimediale) si prevedono le seguenti attività: 1) inventariazione dei beni; 2) studio dello schema di metadati METS e identificazione di un modello descrittivo uniforme per i file digitali; 3) digitalizzazione dei supporti attraverso la strumentazione dell'Università di Macerata secondo gli standard del PND. Questa fase operativa bipartita sarà affiancata da un'azione bilaterale che preveda: 1) studio delle modalità di integrazione dell'AI per il recupero dei file digitali degradati o compromessi; 2) studio delle possibili forme di integrazione dell'AI per la valorizzazione automatizzata dei metadati; 3) riversamento dei contenuti digitali nella Digital Library nazionale (grazie ad apposite convenzioni).

Output (illustrate per ogni attività descritta gli output attesi, specificando tempi e modalità di realizzazione)

1) Collezione cartacea, azione 1 (metadati): presso il CNSL e l'Università di Macerata (Unimc), nel mese 3, studio di METS con output definizione di un modello per la metadattazione dei file digitali immagine; azione 2 (digitalizzazione): presso il CNSL, nei mesi 3-7, tramite strumenti dell'ente, con output produzione di file TIFF per ogni documento. 2) Collezione multimediale, azione 1 (inventariazione): presso il CNSL, nel mese 3, con output descrizione della collezione; azione 2 (metadati): presso il CNSL e l'Unimc, nel mese 4, studio di METS con output definizione di un modello per la metadattazione dei file digitali audio-video; azione 3 (digitalizzazione): presso l'Unimc, nei mesi 4-7, tramite strumentazione d'Ateneo, con output produzione di file digitali per ogni documento. 3) Test e produzione: azione 1 (recovery con AI): presso l'Unimc, nei mesi 3-7, tramite l'infrastruttura di High Performance Computing dell'Ateneo, con output applicazione dell'AI al recupero di file immagine-audio-video compromessi; azione 2 (metadati con AI): presso il CNSL e l'Unimc, nei mesi 8-9, tramite le infrastrutture digitali dei due enti, con output verifica delle possibili forme di integrazione dell'AI nei processi progettuali per la realizzazione dei tracciati dei metadati dei file digitali; azione 3 (riversamento dati): presso il CNSL, nei mesi 9-10, tramite le infrastrutture digitali del CNSL, con output riversamento completo dei dati nella Digital Library nazionale, e loro recupero tramite API per la Biblioteca Digitale Leopardiana. Durante tutto il progetto, entro dieci giorni dalla fine di ciascun mese sarà prodotta documentazione dei processi attuati per poter predisporre relazioni puntuali.

Risultati attesi

I risultati che ci si attendono dal progetto sono di duplice natura. Per quanto riguarda quelli di ricerca: 1) favorire la discussione intorno alla digitalizzazione per formare nuove generazioni di professionisti della cultura in grado di assumere sempre maggiore expertise in un ambito fondamentale per la società contemporanea; 2) incentivare lo studio di tracciati di metadati uniformi per file digitali di varia natura (immagine, audio, video, ecc.) cooperando alla progressiva definizione di standard condivisi e interoperabili con la Digital Library nazionale; 3) incoraggiare l'applicazione virtuosa dell'AI alla valorizzazione dei beni culturali, studiando le modalità con cui sfruttarne le potenzialità per la compilazione dei metadati e il recupero di file digitali degradati o compromessi. Per quanto riguarda i risultati d'azione: 1) completare la digitalizzazione e la metadattazione del fondo Manoscritti del

CNSL (contenente preziosi autografi, come quelli di Giacomo Leopardi); 2) porre le basi per la digitalizzazione/metadattazione del fondo Rari (contenente rarità bibliografiche, tra cui libri con dediche manoscritte di personaggi illustri) selezionando alcuni volumi per un “progetto pilota” che fornirà il modello per il prosieguo delle operazioni; 3) realizzare il censimento, la descrizione, la digitalizzazione e la metadattazione dell’archivio multimediale del CNSL (contenente audiocassette, videocassette e altre testimonianze delle attività promosse dall’ente), muovendo verso la convergenza digitale dei dati bibliografici e archivistici; 4) conferire i dati prodotti alla Digital Library nazionale e recuperare le informazioni tramite API per la Biblioteca Digitale Leopardiana.

Replicabilità delle soluzioni applicate nel progetto (evidenziare gli ambiti di applicazione e replicabilità delle soluzioni e degli output)

Tutte le attività di ricerca-azione previste dal progetto sono pensate per essere replicabili e applicabili sia dal CNSL in iniziative future di valorizzazione digitale delle proprie collezioni, sia da altri istituti culturali che intendano perseguire obiettivi analoghi. Più in particolare: 1) la conduzione di attività di digitalizzazione di supporti cartacei e multimediali consentirà al CNSL di promuovere in autonomia nuove campagne di valorizzazione digitale del patrimonio; al tempo stesso, permetterà agli operatori reclutati di maturare un’importante esperienza applicabile a progetti analoghi sostenuti dal CNSL o da istituzioni diverse; 2) la definizione di un modello di metadati uniforme per la descrizione di file digitali di vario tipo consentirà al CNSL di collaborare alla stesura di un set di metadati amministrativi e gestionali omogeneo a livello nazionale e dunque applicabile a tutti i progetti destinati a confluire nella Digital Library per garantire l’interoperabilità delle soluzioni adottate; 3) lo studio dell’applicazione dell’AI al recupero di file digitali degradati o compromessi, e alla conservazione a lungo termine degli stessi, apre a numerose prospettive nell’ambito della digital curation e favorirà l’avvio di nuove ricerche che consentiranno di formalizzare flussi di lavoro di cui non soltanto il CNSL ma anche altri enti in Italia e nel mondo potranno avvalersi per la gestione e la conservazione delle proprie collezioni digitali; 4) le strategie di applicazione dell’IA alla produzione di metadati potranno essere utili sia al CNSL, sia ad altri, per proseguire l’attività di valorizzazione digitale del proprio patrimonio.

Gruppo di progetto (specificando ruolo, funzioni e competenze interdisciplinari possedute da ciascun gruppo)

Il gruppo di lavoro sarà composto da quattro membri istituzionali e tre reclutati. Appartengono al primo sottogruppo: Gioele Marozzi (coordinatore, umanista digitale: supervisionerà il lavoro promuovendo l’interdipendenza positiva nel gruppo, e curerà la relazione finale di progetto), Laura Melosi (esperta di dominio letterario: garantirà il rigore scientifico del progetto supervisionando le attività di descrizione dei documenti), Stefano Allegrezza (esperto di dominio digitale: supervisionerà le attività di digitalizzazione, di integrazione dell’AI e di metadattazione) e Giovanna Grilli (funzionario amministrativo: si occuperà degli adempimenti economici e della raccolta di documentazione per le relazioni). Appartengono al secondo sottogruppo: un borsista di ricerca ingegnere informatico reclutato dall’Unimc (studierà le forme di integrazione dell’AI agli scopi progettuali), un borsista di ricerca di area archivistica esperto di digitalizzazione reclutato dall’Unimc (curerà lo studio dei metadati relativi alla documentazione multimediale, nonché la digitalizzazione e la metadattazione di quest’ultima e il suo riversamento nella Digital Library nazionale) e un borsista di ricerca di area bibliografico-catalografica esperto di digitalizzazione reclutato dal CNSL (curerà lo studio dei metadati relativi alla documentazione cartacea, nonché la digitalizzazione e la metadattazione di quest’ultima e il suo riversamento nella Digital Library nazionale). Per quanto attiene al piano economico, saranno considerati a valere sul finanziamento di cui al presente bando i costi degli esperti di dominio, della funzionaria amministrativa e dei borsisti. Il coordinatore collaborerà, invece, a titolo non oneroso.

Descrivere sinteticamente l'esperienza maturata dal coordinatore in ambiti legati alla digitalizzazione

Gioele Marozzi è Dottore di Ricerca in Memorie e Digital Humanities. Dal 2017 si occupa a vario titolo di censimento e digitalizzazione di fondi documentali. Dapprima come volontario per il Servizio Civile Nazionale (2017-2018), occupandosi della valorizzazione digitale di documentazione archivistica prodotta dalla famiglia Colocci di Jesi (AN); successivamente durante il Dottorato di Ricerca (2018-2021), occupandosi del censimento, della catalogazione informatizzata, della digitalizzazione (dei supporti e dei testi) e della metadattazione (rispettivamente secondo le codifiche XML-MAG e XML-TEI) dei manoscritti autografi di Giacomo Leopardi conservati in circa 90 enti in Italia e nel mondo; di poi come libero professionista (2021-2022), attraverso collaborazioni di catalogazione del libro antico e moderno secondo gli standard ISBN (A) e (M) e REICAT, nonché mediante la supervisione di un progetto di digitalizzazione realizzato in collaborazione con WikiMedia Italia (produzione di digitalizzazioni e metadati, creazione o adeguamento di pagine Wikipedia, geolocalizzazione su OpenStreetMap); da ultimo come Assegnista di Ricerca (2022-oggi), attraverso un progetto che prevede l'implementazione della Biblioteca Digitale Leopardiana – una piattaforma digitale atta a rendere disponibile tutta la documentazione prodotta intorno a Giacomo Leopardi, partendo dai suoi manoscritti autografi – e il suo arricchimento attraverso la schedatura digitale e la digitalizzazione della Biblioteca e dell'Archivio del CNSL. È stato inoltre docente a contratto presso l'Università di Macerata (a.a. 2023-2024) per gli insegnamenti di “Italianistica e Didattica digitale” e di “Digital History”.

Tipologie di collezioni interessate dall'intervento (almeno due tra museali, archivistiche e bibliografiche) e titolare delle stesse

Le collezioni interessate dall'intervento (archivistica e bibliografica) appartengono al CNSL perché allestite dall'ente o frutto di acquisti diretti / di donazioni; in un caso, vi è in essere una convenzione per deposito e gestione.

Esperienze specifiche nella convergenza digitale (illustrare le esperienze specifiche maturate dal soggetto nell'ambito della convergenza digitale di almeno due delle diverse risorse tra collezioni museali/archivistiche/bibliografiche, con particolare riferimento alla integrazione e valorizzazione delle risorse, in una logica transdisciplinare integrata e inclusiva)

Dal 2017 il CNSL coordina il progetto Biblioteca Digitale Leopardiana, volto a realizzare un collettore digitale in cui offrire al pubblico, non soltanto specialistico, una collezione dedicata a Leopardi e alla sua famiglia, composta dalle schede catalografiche e dalle digitalizzazioni dei manoscritti del poeta, dai testi delle sue opere marcati secondo la codifica XML-TEI, dalle descrizioni dei volumi e dei manoscritti conservati presso la Biblioteca del Centro e dall'inventario delle carte appartenenti all'Archivio dell'Istituto. Lo scopo principale della piattaforma è quello di contribuire alla salvaguardia della memoria leopardiana creando uno strumento gratuito utile a eliminare le distanze che separano documenti e studiosi grazie alla disponibilità in un unico canale di diverse risorse, continuamente aggiornabili a garanzia di scientificità e ricchezza contenutistica. A questo si aggiunge l'impegno che il CNSL sta ponendo nella realizzazione di un percorso espositivo multimediale che abbatta barriere architettoniche e sensoriali per favorire una fruizione inclusiva delle collezioni dell'ente attraverso strategie fondamentali per alcuni ma utili a tutte/i: 1) ascensori per l'accesso; 2) didascalie con QRcode che contengono video parlati (per non vedenti) e sottotitolati (per non udenti) per poter scoprire il patrimonio; 3) copie tattili di beni esposti; 4) un tavolo touchscreen, da cui si potrà accedere anche alla Biblioteca Digitale Leopardiana (favorendo la convergenza digitale tra le tre collezioni del CNSL: bibliografica, archivistica e museale), che potrà essere corredato da un lettore in grado di convertire in Braille quanto visualizzato a schermo per permetterne la fruizione anche ai disabili visivi.

Cronoprogramma (descrivere sinteticamente l'insieme di attività previste ai fini dell'ottimale realizzazione del progetto)

Fase 1 (mesi 1-2): pianificazione, costituzione del gruppo di lavoro, definizione degli obiettivi e avvio delle attività formative

In accordo con l'Unimc, verranno espletate le procedure selettive per titoli e colloquio orale per l'individuazione di due collaboratori (borsisti di ricerca) che saranno incaricati rispettivamente di condurre le attività progettuali di digitalizzazione della collezione multimediale e di studio dell'integrazione dell'AI al recupero di file digitali e alla metadattazione automatizzata degli stessi. Parallelamente il CNSL provvederà al reclutamento di un collaboratore che si occuperà di digitalizzare la collezione cartacea. Contestualmente, il gruppo di lavoro istituzionale si occuperà di definire gli obiettivi del progetto e un cronoprogramma dettagliato delle attività, strutturando anche un protocollo dedicato alle modalità di monitoraggio continuo (ad esempio: produrre documentazione di rendicontazione entro dieci giorni dalla fine di ciascun mese per tenere traccia puntuale di quanto realizzato). Una volta terminate le procedure selettive, il cronoprogramma e tutto quanto stabilito dai membri istituzionali del gruppo di lavoro sarà condiviso con il personale reclutato, che prenderà parte alle attività formative offerte dalla Scuola.

Fase 2 (mesi 3-7): implementazione delle attività principali di ricerca-azione e momenti formativi

Nei mesi 3-7, parallelamente alla partecipazione alle attività formative, il gruppo di lavoro avvierà le attività di ricerca-azione, secondo una logica incrementale volta a garantire la scientificità, la scalabilità e la replicabilità dei risultati. Collaborando con il coordinatore e sotto la supervisione scientifica degli esperti di dominio: 1) collaboratore CNSL, mese 3: studierà lo standard METS per i file digitali immagine; mesi 3-7: digitalizzerà il fondo Manoscritti e a una selezione di libri del fondo Rari del CNSL. 2) borsista Unimc n. 1, mese 3: si occuperà dell'inventariazione e della descrizione della collezione multimediale del CNSL; mese 4: studierà lo standard METS per i file digitali audio-video; mesi 4-7: digitalizzerà l'Archivio multimediale. 3) borsista Unimc n. 2, mesi 3-7: affiancherà il collaboratore CNSL e l'altro borsista Unimc nell'analisi dello standard METS per testare l'implementazione dell'AI alla valorizzazione automatizzata dei metadati, e studierà strategie di sfruttamento dell'AI per il recupero dei file immagine, audio o video degradati o compromessi, e per favorire la conservazione a lungo termine dei file digitali.

Fase 3 (mese 5): analisi dello stato di avanzamento e adeguamento del progetto, se necessario

In questa fase, anche grazie alla supervisione del tutor e della Scuola, si verificherà l'aderenza del progetto al cronoprogramma, cercando di rispondere alle eventuali criticità riscontrate per risolverle definitivamente. Il collaboratore CNSL e entrambi i borsisti Unimc elaboreranno una relazione completa e dettagliata dei primi mesi di lavoro, in cui daranno contezza quantitativa e qualitativa del proprio lavoro, dei traguardi raggiunti, delle criticità riscontrate e delle eventuali difficoltà ancora da affrontare. Si proseguirà inoltre normalmente con le attività di digitalizzazione dei documenti cartacei e audio-visivi, nonché con quelle di studio delle forme di integrazione dell'AI alla valorizzazione dei metadati e al recupero di file digitali compromessi. In questo mese sono attesi i primi test ufficiali dell'implementazione dell'AI per verificare la bontà della ricerca in atto e inserire eventuali correttivi. Contestualmente, il funzionario amministrativo collaborerà alla stesura di una prima fase ufficiale di rendicontazione progettuale, da condividere con il tutor e la Scuola.

Fase 4 (mesi 8-10): conclusione delle attività, valutazione finale e produzione degli output

In questa fase le azioni del gruppo di lavoro convergeranno verso il raggiungimento degli output e dei risultati progettuali. Previo confronto con il tutor e con la Scuola, si provvederà all'approvazione del modello uniforme di metadattazione studiato dai borsisti Unimc e dal collaboratore CNSL; in seguito, si realizzeranno nuovi test di implementazione dell'AI per la metadattazione automatizzata dei file digitali, così da verificare la bontà del processo di metadattazione degli stessi. Da ultimo, si provvederà, attraverso le infrastrutture ministeriali, al conferimento alla Digital Library nazionale di

tutti i file realizzati e dei relativi metadati, nonché al loro recupero tramite API per la Biblioteca Digitale Leopardiana. Nel corso del mese 10, inoltre, sotto la supervisione di entrambi gli esperti di dominio e del coordinatore, saranno valutati definitivamente i risultati raggiunti, e sarà stilata la relazione finale del progetto, in cui saranno riportati – per garantirne la scientificità, la misurabilità e la replicabilità – tutti i processi attuati, gli output prodotti, i modelli utilizzati e le opportunità di ricerca aperte.