DESCRIZIONE DEL PROGETTO:

a) Titolo del progetto:

Intelligenza artificiale e Memoriali bolognesi:

potenzialità e limiti della trascrizione automatica per i corpora documentari medievali

b) Acronimo del progetto:

IAMBO

c) Indicazione di massimo 3 settori ERC (in ordine di pertinenza) per il progetto:

SH6\_7: Medieval history

SH6\_5: Palaeography and codicology

SH6\_12: Social and economic history

d) Indicazione di massimo 5 parole chiave

Fonti notarili seriali;

Paleografia;

Storia sociale;

Digital humanities;

HTR (Handwritten Text Recognition)

e) Descrizione degli **obiettivi** del progetto di ricerca (massimo 4.000 caratteri)

IAMBO intende testare le potenzialità delle tecnologie di riconoscimento automatico di scrittura (*Handwritten Text Recognition* = HTR) per lo studio delle fonti notarili seriali tardomedievali, una tipologia di corpora documentari di dimensioni monumentali che li hanno resi, finora, elusivi agli sforzi di una ricerca sistematica.

HTR è una delle tecnologie *machine-learning* più promettenti nel campo delle *digital humanities*, perché in grado di dare accesso ai contenuti di alcune fonti manoscritte riducendo, in maniera sostanziale, l’impegno umano per la trascrizione. La tecnologia HTR è in fase matura, tale da potere affrontare problemi specifici (dal riconoscimento di layout complessi alla riduzione della *ground-truth* necessaria per allenare un modello). Inoltre, lo sviluppo di piattaforme come Transkribus risulta assai promettente nel campo storico, perché rende utilizzabile la HTR anche senza conoscenze informatiche approfondite. In questo quadro, le scritture notarili, a causa del grado di corsività e del numero di abbreviazioni, restano tra le più complesse e maggiormente bisognose di ulteriori test, operazione per la quale sono indispensabili conoscenze paleografiche e diplomatistiche. Allo stesso tempo, chi ha competenze storiche è chiamato a misurarsi, in chiave anche critica, con le possibilità offerte dalla HTR rispetto a un tipo di fonte assai ricco per indagare la vita economica e sociale del tardo medioevo.

Nel concreto ci si concentrerà sui registri pubblici prodotti dall’ufficio dei Memoriali del Comune di Bologna, istituito nel 1265 e attivo fino al 1452 con lo scopo di garantire la certezza dei diritti dei privati: tutti gli atti stipulati, in città e nel contado, superiori alle 20 lire di bolognini erano registrati all’interno dei registri, compilati da notai. Da qui è scaturita un’eccezionale serialità documentaria, molto nota agli studiosi ma analizzata solo attraverso carotaggi selettivi, vista la sua mole (oltre 2400 registri, rilegati in 322 volumi). Oltre che per la sua rilevanza, la scelta di questa fonte documentaria permetterà di sfruttare l’esperienza accumulata nel più ampio progetto di ricerca MemoBo, attivo presso DISCI dal 2019.

In questo quadro, l’obiettivo specifico di IAMBO è: creare un modello di lettura per tali documenti; confrontarsi con le difficoltà (pratiche e metodologiche) che l’operazione pone; misurare la base di dati (i.e. trascrizione di documenti) necessaria per allenare un modello e valutarne il grado di affidabilità. Dal punto di vista tecnologico, il progetto sarà sviluppato in collaborazione con Regesta, azienda specializzata in *New Media and Historical Heritage*, già partner di MemoBo, la qualesi è già detta interessata a supportare la ricerca legata alla HTR. Per testare la capacità di HTR, saranno indispensabili competenze paleografiche e diplomatistiche, sia per il necessario controllo, sia per riflettere sulle connesse questioni metodologiche. In tale prospettiva, l’assegnista in IAMBO svilupperà anche una linea di ricerca che serva a mettere in luce quello che l’intelligenza artificiale da sola non può fare, cioè un’analisi storica dei dati, da concretizzarsi in un articolo.

L’obiettivo ultimo di IAMBO è la verifica delle possibilità (e/o limiti) della HTR per i Memoriali bolognesi, passaggio nodale per il macro-progetto MemoBo, su cui DISCI ha investito in questi anni. Un successo nell’impiego della HTR costituirebbe un decisivo salto di qualità, trasformando e accelerando la raccolta dei dati, là dove l’esperienza maturata in IAMBO fornirebbe un modello per la loro elaborazione. Inoltre, la competenza accumulata nell'uso di tali strumenti per la ricerca storica fornirà la base per futuri lavori su fonti affini. L’emergere di limiti o problemi nella HTR sarà ugualmente prezioso, perché permetterà una nitida prospettiva su quali siano le piste attualmente percorribili nello studio delle fonti documentarie seriali.

f) Descrizione del carattere di **originalità e innovatività** della proposta progettuale (massimo 3.000 caratteri);

L’aspetto innovativo del progetto si lega, come già illustrato, allo sperimentare le potenzialità della HTR come tecnologia *machine-learning* riguardo a una tipologia documentaria ampia e di particolare complessità (dal punto di vista paleografico, diplomatistico e storico) come sono i registri dell’ufficio dei Memoriali. Verificare la possibilità di istruire la macchina per trascrive le fonti notarili tardomedievali – e individuare i passaggi necessari e le criticità di questo processo – rappresenta il passaggio imprescindibile per fare un salto di qualità nello studio sistematico del vastissimo corpus documentario rappresentato da questo fondo archivistico che, stando a una stima affidabile, contiene la registrazione di più di un milione di contratti di varia natura (mutui, testamenti, vendite, doti…), stipulati tra privati a Bologna e nel suo contado nell’arco di poco meno di due secoli. In tale senso, si tratta di fonti attraverso cui la città rivela significativi tratti del suo profilo, fornendo (involontariamente) una sorta di propria carta di identità. Non sembra quindi inopportuno usare per questo progetto l’acronimo IAMBO, leggibile anche come: I am Bo.

Calibrare la tecnologia HTR su questa tipologia di documenti mira ad aprire una fase davvero nuova nello studio di quello che è stato definito «un mare magnum di possibilità», ovvero una fonte che – vista la sua natura organica e continuativa – permette di leggere in controluce una vastissima gamma di fenomeni sociali, economici, culturali. In questa prospettiva, in stretto dialogo con Regesta.exe, MemoBo (incentrato su una schedatura completa di questa fonte) ha individuato la necessità e l’opportunità di testare la tecnologia HTR come snodo decisivo e più promettente per fare crescere il progetto avviato. IAMBO si inserisce organicamente in tale quadro, quale passaggio insieme necessario e innovativo. Inoltre, IAMBO intende affrontare la sperimentazione della tecnologia HTR come questione non meramente tecnica, ma come occasione di riflessione critica sul contributo che le nuove tecnologie forniscono (o possono fornire) tanto allo studio del medioevo quanto alla paleografia e alla diplomatica. L’impatto del progetto IAMBO, come sarà delineato meglio sotto (punto i), sarà quindi duplice: diretto e misurabile sullo sviluppo del progetto MemoBo; indiretto e potenziale su futuri progetti che prendano in considerazione corpora di fonti seriali (tardomedievali) analoghe.

g) Linea di intervento A: Descrizione della **coerenza con i temi di rilievo trasversale per il PNR 2021**- ALMA IDEA 2022 7 2027 (massimo 3.000 caratteri)

Il progetto IAMBO risponde ad alcune priorità di sistema individuate nel PNR 2021-27. Anzitutto, “Rafforzare la ricerca interdisciplinare”. Il progetto favorisce la consapevole contaminazione di competenze (storiche, paleografiche, digitali), riunendo approcci scientifici diversi sia per affrontare una sfida tecnica, quella della HTR applicata alle fonti medievali, sia per costruire la base di ricerche future su molteplici livelli, a partire da una diretta applicazione nel progetto MemoBo. Inoltre, sviluppando un dialogo tra università, istituzioni pubbliche (Archivio di Stato di Bologna) e professionisti (Regesta), IAMBO contribuisce a formare uno spazio dinamico per la ricerca, come chiesto dalle linee guida del PNR.

IAMBO si collega inoltre direttamente alla priorità di “Accompagnare lo sviluppo di una nuova generazione di ricercatori e professionisti del trasferimento di conoscenza”. Chi lavorerà nel progetto maturerà infatti preziose competenze riguardo allo sviluppo e gestione di *dataset* per la ricerca storica e alle problematiche e potenzialità relative al renderli accessibili come open data. Una specifica e immediata ricaduta dei risultati di IAMBO, inoltre, si potrà avere nell’offerta formativa, integrandoli nel *Laboratorio di medievistica*, già attivo presso DISCI, dedicato a formare studenti/esse riguardo al “mestiere dello storico" in rapporto alle nuove tecnologie, e all’attività didattica per dottorandi e studenti magistrali in programma presso FICLIT. La sperimentazione dell’applicazione HTR si inserisce infatti pienamente nelle finalità di questi ambiti didattici. In tale senso, il progetto risponde anche a un’altra priorità del PNR, là dove parlando del “Consolidare la ricerca fondamentale” invita a perseguire una relazione stretta tra ricerca fondamentale ed educazione.

Infine, IAMBO si connette alla priorità di “Valorizzare la circolazione di conoscenza e competenze”, non solo – come accennato – per la collaborazione con enti pubblici e con partner privati (incontro tra ricercatori e imprese), ma perché i dati raccolti, confluendo nel database di MemoBo, rientrano nelle pratiche di *open innovation* e scienza aperta. Il database è infatti pensato per produrre e gestire *linked open data* (LOD), in modo conforme alla politica di dati FAIR (*Findable, Accessible, Interoperable, Reusable*), sostenuta nel PNR. In tal senso, il progetto parteciperà alla creazione collettiva dei «big data of the past», obiettivo che anima diversi progetti di interesse nazionale ed europeo e che, in ultima analisi, mira a favorire la fruibilità su larga scala della cultura e della ricerca scientifica, aprendo a tutti gli esiti della ricerca in modo ampio e inclusivo. In tal senso, pure nella sua specificità, IAMBO guarda a una delle frontiere più vivaci della ricerca, collocandosi all’intersezione di due ambiti di intervento individuati dal PNR, quello relativo alla Cultura umanistica e quello relativo al Digitale e alla Intelligenza artificiale.

h) Illustrazione dei profili di **interdisciplinarità della ricerca e complementarietà di competenze tra PI e CO-PI** (massimo 3.000 caratteri)

Come delineato nelle precedenti sezioni, la ricerca di IAMBO ha una spiccata natura interdisciplinare, perché richiede il dialogo e la collaborazione tra abilità paleografiche e diplomatistiche (in particolare, legate al notariato medievale), la capacità di una lettura in chiave storica dei dati emersi (con particolare attenzione alla dimensione sociale ed economica delle città tardomedievali), il padroneggiare le tecnologie delle *digital humanities*, legate anche alla costruzione del cosiddetto web-semantico.

IAMBO si avvarrà in questo dell’esperienza maturata nel progetto MemoBo, attivo presso DISCI e all’interno del quale Pietro Delcorno (PI) e Maddalena Modesti (CO-PI) hanno già avuto modo non solo di collaborare, ma di pensare – in dialogo con altri membri di MemoBo (in particolare, Tommaso Duranti e, per Regesta, Giovanni Bruno) – le linee guida e le finalità di IAMBO.

Nello sviluppo del progetto, l’assegnista di ricerca, che dovrà avere spiccate competenze medievistiche e paleografiche, potrà contare sulle competenze specifiche di PI e CO-PI. Delcorno (PI) metterà a disposizione di IAMBO le sue conoscenze riguardo alla società urbana tardomedievale e ad alcuni aspetti di storia economica (quali quelli legati alle reti formali e informali del credito), nonché le competenze e riflessioni metodologiche maturate in precedenti progetti dotati di una componente di *digital humanities* (in particolare, *Lenten sermons*, di cui è PI e che ha incluso la costruzione di un database). Modesti (CO-PI) garantirà allo sviluppo del progetto l’esperienza necessaria nel campo della trascrizione e analisi, paleografica e diplomatistica, di documenti notarili medievali e un’estesa conoscenza dei materiali dell’Archivio di Stato di Bologna e di altri fondi documentari, maturata come responsabile scientifico del centro studi RAM (Ricerche e Analisi Manoscritti) e di ADLAB (Analogico-Digitale Laboratorio) di FICLIT. Le competenze di entrambi saranno utili anche a immaginare future applicazione della tecnologia HTR.

La complementarietà tra PI e CO-PI permetterà, in un certo senso, uno sguardo sia interno (paleografico) che esterno (storiografico) ai documenti presi in analisi, favorendo la collaborazione interdipartimentale tra DISCI e FICLIT, già sperimentata in MemoBo e nel centro RAM. IAMBO permetterà quindi di rafforzare, in maniera fattiva, tale collaborazione, offrendo inoltre la possibilità di immaginare sia progetti di ricerca più estesi (vedi punto j), sia collaborazioni dal punto di vista didattico (oltre a quanto delineato al punto i) su temi trasversali come la lettura, l’interpretazione, e la valorizzazione della documentazione archivistica tardomedievale, settore strategico per entrambe le discipline (M-STO/01 e M-STO/09) e – insieme – in forte trasformazione alla luce dalle nuove tecnologie digitali, rispetto alle quali è proficuo che DISCI e FICLIT sviluppino esperienze specifiche e condivise.

i) Linea di intervento A: Illustrazione dell’i**mpatto** della ricerca (max 3.000 caratteri)

L’impatto di IAMBO – come già richiamato – si declina in tre direzioni: lo sviluppo del progetto MemoBo; le potenzialità per futuri progetti su fonti comparabili; l’offerta formativa.

1) Oltre al salto di qualità che la tecnologia HTR può apportare a MemoBo (in dettaglio sopra, punto f), va qui richiamato come i dati raccolti da IAMBO verranno riversati nel database digitale di MemoBo, costruito con codici *open source* sulla piattaforma xDams, grazie alla collaborazione con Regesta. Il database, disegnato per le necessità specifiche della schedatura dei Memoriali, è ora in versione beta e *backend*, ma destinato a essere presto aperto al libero accesso e utilizzo online. L’impatto di IAMBO sulla crescita di questo database è duplice: diretto, arricchendo il *dataset* e favorendo la messa a punto dello strumento; indiretto, testando modalità e possibilità per accelerare la raccolta dei dati con HTR e mettendo in luce il tipo di lavoro, paleografico e storiografico, richiesto da questo processo. Cooperando con lo sviluppo del database di MemoBo (progetto su cui DISCI ha scommesso e investito energie e risorse in questi anni), IAMBO contribuirà inoltre alla riflessione e alle iniziative legate al vasto tema della migrazione digitale del patrimonio culturale, inteso come processo non limitabile alla creazione di copie digitali dei corrispettivi oggetti analogici, ma che implica un ripensamento delle pratiche di accesso, produzione, circolazione dei contenuti culturali digitali, così come di quelle della ricerca storica, alla luce dei nuovi contesti tecnologici (segnati da grandi vantaggi e non minori problematicità).

2) L’esperienza accumulata con IAMBO su possibilità e limiti di HTR per i Memoriali potrà essere trasferita a future indagini su fonti affini, aprendo la strada a ricerche quantitative e qualitative su larga scala che offrono anche la possibilità di creare un *corpus* di dati omogenei (confrontabili e navigabili) per la ricerca storica medievistica. In tal direzione si muove già MemoBo e a tale sforzo corale contribuirà IAMBO, favorendo un ampliamento delle *open* *libraries* semantiche necessarie al *machine-learning*, in condivisione con altri gruppi di ricerca nazionali e internazionali, seguendo una strategia collaborativa indispensabile per raggiungere progressi in questo campo.

3) IAMBO avrà inoltre un impatto nell’offerta formativa (DISCI/FICLIT). L’esperienza acquisita riguardo all’uso di HTR per le fonti tardomedievali favorirà la riflessione (anche metodologica) su tale tema con studenti/esse che si formano nelle scienze storiche e paleografiche. Seguendo una prassi già sperimentata con MemoBo, sarà cura di PI e CO-PI garantire una proficua interazione tra ricerca e didattica, tenendo conto di come le competenze nell’ambito delle *digital humanities* siano sempre più richieste in chiave professionale per affiancare tecnici e amministratori nei processi di transizione digitale nel settore della cultura e dei beni culturali.

j) Illustrazione **dell’implementazione** della ricerca (attività, work plan, ecc.), con particolare riferimento alla **fattibilità e sostenibilità** in relazione al finanziamento erogato (massimo 6.000 caratteri).

**Work plan**

Cuore di IAMBO sarà il lavoro svolto dal/la titolare dell’assegno di ricerca annuale (inizio: settembre 2022 circa), che si svilupperà in tre fasi, precedute da una fase preparatoria e una post-project.

a) Fase preparatoria (2 mesi), a cura di PI e CO-PI:

* Ricognizione dei materiali necessari per l’avvio del progetto (in particolare, fotoriproduzioni e trascrizioni già in possesso di MemoBo).
* Coordinamento con Regesta per precisare le fasi di lavoro sulla HTR riguardo al registro dei Memoriali del 1287 individuato come base per un primo test.
* Bando di selezione dell’assegnista.

b) Prima fase assegno (4 mesi)

* L’obiettivo principale sarà testare (in collaborazione con Regesta) la tecnologia HTR sul registro dei Memoriali del 1287, individuato come *ground-truth* per allenare un modello. In concreto si tratterà di calibrare il software grazie al confronto tra trascrizioni e riproduzioni digitali (già a disposizione grazie a MemoBo).
* Verifica dei dati relativi a questo registro già inseriti nel database di MemoBo.
* Valutazione – in collaborazione con PI, CO-PI e personale di Regesta – di potenzialità, limiti e problemi da risolvere riguardo all’uso della HTR (*first assesment*). PI e CO-PI decideranno se coinvolgere in questo processo altre persone di MemoBo.

c) Seconda fase assegno (4 mesi)

* Individuare un secondo registro dei Memoriali che: a) rispecchi gli interessi di ricerca del/la assegnista; b) sia funzionale come base per un secondo test di HTR, alla luce di quanto emerso nella fase precedente.
* Trascrivere il registro e schedarlo nel database di MemoBo (con l’obiettivo intermedio di accrescerne il *dataset*). L’inserimento dei dati potrà estendersi anche alla fase successiva.
* Individuare una linea di indagine nei documenti presi in considerazione, discutendo con PI e CO-PI l’abstract di un articolo sul tema (prodotto della ricerca conferibile nell’ambito VQR, da pubblicare entro un anno).
* *Work-in-progress seminar* sullo stato della ricerca IAMBO, coinvolgendo *in primis* i membri di MemoBO.

d) Terza fase assegno (4 mesi)

* Seconda fase di test (collaborazione con Regesta) della tecnologia HTR sul nuovo registro, individuato come *ground-truth*.
* Valutazione del secondo test e *second assesment* (collettivo) di potenzialità e limiti nell’utilizzo di HTR, da declinare sia per il progetto MemoBo, sia per documentazione simile. Il risultato della riflessione sarà condensato in un protocollo utile per gli sviluppi successivi.
* Stesura e invio dell’articolo sul tema prescelto. In accordo con PI e CO-PI (e in linea con il bando Alma Idea) tale impegno potrà essere posticipato nel caso in cui si giudichi più opportuno concentrarsi sulla scrittura di un bando per ricerca competitiva finalizzato a proseguire e ampliare il progetto (cfr. sotto).
* Presentazione da parte di assegnista, PI e CO-PI dei risultati di IAMBO e delle prospettive che si aprono a breve/medio termine (presso DISCI o Archivio di Stato di Bologna).

e) Fase post-project (3 mesi), a opera di PI e CO-PI, con il coinvolgimento di chi ha terminato l’assegno.

* Rendicontazione scientifica e finanziaria.
* Insieme ai membri di MemoBo, valutazione di come i risultati di IAMBO incidano sullo sviluppo di tale progetto.
* Nel caso la sperimentazione della HTR si confermi proficua per lo studio delle fonti manoscritte seriali, individuare fonti di finanziamento (nazionali e europee) per un più esteso sviluppo del progetto (IAMBO2).

Attività trasversali al lavoro del/la assegnista (fasi b, c, d):

Raccordando lo sviluppo di IAMBO e il calendario didattico, l’assegnista sarà coinvolto/a sia nel laboratorio di medievistica (DISCI), sia nel corso di paleografia e diplomatica e nell’attività didattica di ADLAB (FICLIT), in accordo con i rispettivi docenti e referenti scientifici (già dettisi interessati a integrare ricerca e didattica).

L’assegnista dovrà tenere, nelle forme da lui/lei scelte, una sorta di ‘diario di bordo’ del proprio lavoro, con particolare attenzione agli aspetti metodologici e alla riflessione critica su come le tecnologie incidano sulla ricerca storica, in modo da poter fornire – in forme da definirsi (e.g. micro-video nel portale MemoBo) – uno sguardo personale e riflessivo sul progetto IAMBO.

**Fattibilità e sostenibilità**

Fondandosi sull’esperienza e i materiali accumulati nella fase pilota di MemoBO, IAMBO si trova in una posizione privilegiata che permette di concentrare immediatamente le forze su un obiettivo specifico, avendo a disposizione un’eccezionale base di partenza costituita dalle trascrizioni già effettuate, dalle riproduzioni dei documenti, dalla presenza di un database specifico in fase avanzata di sperimentazione. Costruendo su questa esperienza concreta, IAMBO ha il notevole vantaggio di avere un quadro preciso dei passaggi da compiere e delle loro tempistiche, elemento dirimente per un progetto breve come Alma Idea. In particolare, la presenza del database permette l’uso (e la valorizzazione) di uno strumento raffinato, senza costi aggiuntivi.

Va inoltre sottolineata la favorevole congiuntura rappresentata dalla collaborazione con Regesta, partner professionale con competenze specifiche nella costruzione di archivi digitali sul web. L’interesse di Regesta per questo progetto e la sua volontà di supportarlo con le proprie competenze non è ovvia e testimonia come, da esperti nel settore, si riconosca – a MemoBo e allo sviluppo consentito da IAMBO – un potenziale innovativo e un forte valore culturale.

Da ultimo, il fatto che IAMBO contribuisca a un progetto già impiantato garantisce che i risultati abbiano un impatto immediato, non effimero o legato al solo percorso del/la assegnista. In tal senso, l’interscambio tra MemoBo e IAMBO si presenta come un elemento di forza del progetto, inteso a valorizzare, rafforzare e spostare verso nuove sfide il lavoro fatto negli anni precedenti, frutto di una fattiva collaborazione interdisciplinare che è già stata capace di coinvolgere docenti, dottorandi e studenti.