



LACUNA – “Leveraging innovative Approaches to Comprehensively UNderstand Ancient Epicurean Texts. Towards the First AI-Enhanced Editions of Herculaneum Theological Papyri”

progetto finanziato nell'ambito del Fondo italiano per la Scienza 2022–2023 (Bando FIS 2).

Codice progetto: FIS-2023-01833

CUP: E53C24003870001 CUI: S00876220633202500005

CAPITOLATO TECNICO DELLA GARA D'APPALTO

INDICE

ART. 1 – OGGETTO DELL'APPALTO

ART. 2 – CARATTERISTICHE TECNICHE E FUNZIONALI DEL SOFTWARE RICHIESTO

ART. 3. – PIANO DI LAVORO E TEMPI DI REALIZZAZIONE DEL PROGETTO

ART. 4 – MODALITÀ DI EROGAZIONE DEL SERVIZIO E MODALITÀ DI PAGAMENTO

ART. 5. – OBBLIGO DI CONSEGNA DEI DATI NEL CORSO DEL PROGETTO

ART. 6. – CLAUSOLA DI RISOLUZIONE E GARANZIE

ART. 7. – PROPRIETÀ INTELLETTUALE

ART. 1. OGGETTO DELL'APPALTO

Il presente appalto ha per oggetto l'attività di ricerca volta alla progettazione e allo sviluppo di Theion, un software basato sull'applicazione di modelli di intelligenza artificiale alla ricostruzione testuale dei papiri ercolanesi, da realizzarsi in un periodo di tre anni, da parte di una compagnia esterna specializzata. L'obiettivo finale del progetto è la consegna di un software completo e funzionante, conforme alle specifiche tecniche definite dall'Host Institution nell'Art. 3.

ART 2. CARATTERISTICHE GENERALI DEL SOFTWARE RICHIESTO

Il software Theion dovrà essere sviluppato nell'ambito del progetto triennale LACUNA e rappresenterà una soluzione avanzata per la ricostruzione testuale dei papiri greci ercolanesi frammentari, precisamente di quelli che restituiscono le opere teologiche di Filodemo di Gadara oggetto del Progetto. Secondo il piano di lavoro indicato nel successivo Art. 3, il software dovrà implementare modelli di AI per:

1. fare predizioni linguistiche per suggerire integrazioni di lacuna sulla base del contesto e dello stile dell'autore. I modelli di AI saranno allenati sul *corpus* dei testi in greco antico e, nello specifico, sul *corpus* dei testi di Filodemo di Gadara, in parte già disponibili su *papyri.info*, in parte appositamente implementati nel corso del Progetto. I modelli di AI impiegati per la previsione dei testi mancanti dovranno partire sia da input umani che dalle immagini dei papiri ercolanesi. Il software, dunque, dovrà integrare metodologie di AI per il riconoscimento automatico dei caratteri nelle immagini dei papiri e per la segmentazione delle sequenze di testo, con output in forma di trascrizione diplomatica.
2. realizzare *text mining* automatizzato per suggerire paralleli testuali e contenutistici pertinenti basati su criteri di similarità linguistica e concettuale.

ART. 3. PIANO DI LAVORO E TEMPI DI REALIZZAZIONE DEL PROGETTO

L'oggetto della commessa di ricerca verrà svolto attraverso quattro Milestones distribuite sui 3 anni del Progetto:

MILESTONE 1 (ANNO 1): Raccolta dati per l'addestramento AI; HTR sulle immagini dei papiri

Semestre 1. Raccolta dati per l'addestramento AI. Nel primo semestre del Progetto è prevista la ricerca e la raccolta di set di dati appropriati per l'addestramento di modelli di intelligenza artificiale in grado di fare previsioni linguistiche per l'analisi e la ricostruzione dei testi di Filodemo attestati in papiri ercolanesi greci frammentari. Dal punto di vista metodologico, ciò includerà l'accesso a repository open-source del *corpus* di tutti i testi greci antichi disponibili. Per consentire ai modelli di intelligenza artificiale di riconoscere lo stile, la sintassi e il vocabolario peculiari di Filodemo, si richiede che questi modelli siano addestrati sul *corpus* digitale completo di opere note dell'autore disponibile su *papyri.info*, che verrà ulteriormente incrementato con nuovi testi nel corso del Progetto.

Semestre 2. HTR sulle immagini dei papiri. Poiché si richiede che i modelli di AI impiegati per la previsione dei testi mancanti partano sia da input umani che dalle immagini dei papiri ercolanesi, nel secondo semestre del progetto il software dovrà integrare modelli di intelligenza artificiale per il riconoscimento automatico dei caratteri (HTR) nelle immagini multispettrali dei papiri selezionati, oltre a segmentare le sequenze di testo, producendo un output in forma di trascrizione diplomatica. Sfruttando l'applicazione dei modelli AI alle immagini, sarà necessario cominciare a investigare la possibilità di offrire, nell'interfaccia utente di Theion, una visualizzazione fotorealistica delle ricostruzioni testuali nelle lacune, da realizzare nell'anno 2. Merita, inoltre, un'indagine approfondita la sinergia tra i modelli AI per l'integrazione delle lacune e i sistemi HTR applicati ai papiri con stratigrafia complessa, al fine di migliorare la ricostruzione di sequenze testuali interrotte.

MILESTONE 2 (ANNO 2): Addestramento dei modelli AI; sviluppo di Theion (fase Alpha)

Il secondo anno del progetto sarà dedicato a due attività organizzate in due semestri:

Semestre 1. Addestramento dei modelli AI per l'integrazione delle lacune e per il text mining dei paralleli lessicali e contenutistici.

- Integrazione delle lacune: per verificare l'accuratezza predittiva dei modelli dovranno essere vagliati gli approcci più adatti, tra cui: confronto tra testi con lacune reali e testi privi di lacune (creazione di fake gaps nei testi completi per testare la capacità del modello di prevedere i segmenti mancanti); confronto tra le integrazioni proposte dall'AI e le integrazioni umane. È fondamentale in questa fase l'interazione tra il team della compagnia appaltatrice e il team papirologico del Progetto.
- Text mining dei paralleli: i modelli AI applicheranno text mining automatizzato per suggerire paralleli testuali e contenutistici pertinenti che aiuteranno lo studioso nel processo di edizione del testo. L'individuazione dei paralleli non dovrà basarsi su una corrispondenza parola per parola (come già è possibile fare sfruttando le funzioni di "Text search" e "Wildcard search" del TLG), ma su criteri di similarità linguistica, stilistica e concettuale.

Semestre 2. Sviluppo di Theion (fase Alpha):

Creazione del prototipo con le funzionalità fondamentali, focalizzato sulla ricostruzione delle lacune e sulla gestione delle alternative. In questa fase, Theion dovrebbe:

- avere un'infrastruttura software di base che consenta il caricamento di testi e immagini;
- offrire una visualizzazione integrata in cui sono affiancate: l'immagine multispettrale del papiro, la trascrizione (inserita dallo studioso e/o generata dall'HTR), il testo integrato proposto dall'AI;
- fare predizione automatica delle integrazioni sulla base dello spazio disponibile nella lacuna, della lingua greca, del contesto del passo, dello stile dell'autore, dei paralleli individuati;
- mostrare le alternative classificate in base alla loro probabilità linguistica e contestuale;

- indicare la probabilità di correttezza delle proposte con una valutazione metrica;
- offrire una visualizzazione fotorealistica delle ricostruzioni testuali nelle lacune sulla base dei caratteri identificati con l'HTR e con highlights dinamici (selezionando una lacuna nel testo si evidenzia la zona dell'immagine corrispondente, e viceversa);
- permettere all'utente di segnalare errori o migliorare le proposte tramite feedback;
- possibilità di inserire manualmente integrazioni e confrontarle con quelle AI;
- memorizzare le diverse ipotesi ricostruttive;
- avere un database per l'archiviazione dei dati.

MILESTONE 3 (ANNO 3): Sviluppo di Theion (fase Beta)

Semestre 1. Completamento delle funzionalità avanzate di Theion:

- Fine-tuning delle funzioni sviluppate nella fase Alpha;
- ottimizzazione dell'algoritmo AI per affinare l'integrazione delle lacune;
- miglioramento delle spiegazioni sulle alternative: giustificare perché una ricostruzione è preferibile mostrando paralleli testuali e contenutistici individuati tramite text mining;
- modulo di ricerca paralleli: offrire la possibilità di escludere determinate fonti o filtrarle in base alle preferenze;
- Link diretto a paralleli testuali (da TLG, papyri.info, ecc.)
- interfaccia più avanzata e personalizzabile;
- possibilità di esportazione dei dati.

Semestre 2. Rifinitura, testing, e finalizzazione di Theion per il rilascio al pubblico:

- Rifinitura dell'interfaccia utente;
- Integrazione del feedback degli utenti.

MILESTONE 4 (ANNI 1-3): Diffusione dei risultati

- anni 1-3: incontri mensili tra il team della società appaltatrice, il PI e il team papirologico del Progetto;
- anni 1-3: presentazione di articoli su riviste nei campi della papirologia, dell'informatica e/o delle Digital Humanities;
- fine anno 2/inizio anno 3: workshop di formazione su Theion;
- entro la conclusione del progetto: elaborazione di un MOOC (Massive Open Online Course) con istruzioni e video tutorial sull'utilizzo di Theion;
- partecipazione all'evento conclusivo del Progetto.

ART. 4. MODALITÀ DI EROGAZIONE DEL SERVIZIO E MODALITÀ DI PAGAMENTO

La compagnia aggiudicataria dovrà garantire lo sviluppo progressivo del software attraverso rilasci periodici. Il pagamento della somma aggiudicata rispetto alla somma a base di gara (pari a 300.000 € **oltre IVA**) sarà suddiviso in sei tranches semestrali, pari a 1/6 della somma aggiudicata, che verranno erogate all'espletamento delle attività descritte nell'art. 3 secondo le seguenti condizioni.

Anno 1

Semestre 1. Raccolta dati per l'addestramento AI (pagamento I tranche)

La società appaltatrice dovrà fornire un elenco completo e dettagliato dei dataset raccolti per l'addestramento dei modelli AI.

Semestre 2. HTR sulle immagini dei papiri (pagamento II tranche)

La società appaltatrice dovrà fornire:

- Software funzionante di HTR applicato alle immagini multispettrali dei papiri selezionati.
- Campioni di immagini annotate con trascrizioni diplomatiche ottenute tramite il sistema HTR.

Anno 2

Semestre 1. Addestramento dei modelli AI per integrazione lacune e text mining (pagamento III tranche)

La società appaltatrice dovrà fornire:

- Modelli AI addestrati per l'integrazione delle lacune e il text mining sul *corpus* digitale completo di opere note di Filodemo;
- Evidenza di test svolti su testi con lacune reali e "fake gaps";
- Confronto tra output AI e integrazioni umane (con documentazione di metodo e risultati);
- Risultati documentati del *text mining* automatizzato con esempi di paralleli identificati.

Semestre 2. Sviluppo di Theion fase Alpha (pagamento IV tranche)

La società appaltatrice dovrà fornire un prototipo funzionante di Theion che includa tutte le funzioni descritte nell'Art. 3, Milestone 2 semestre 2, cioè:

- Infrastruttura software di base che consenta il caricamento di testi e immagini;
- Visualizzazione integrata in cui sono affiancate: l'immagine multispettrale del papiro, la trascrizione (inserita dallo studioso e/o generata dall'HTR), il testo integrato proposto dall'AI;
- Capacità di predizione automatica delle integrazioni sulla base dello spazio disponibile nella lacuna, della lingua greca, del contesto del passo, dello stile dell'autore, dei paralleli individuati;
- Visualizzazione delle alternative classificate in base alla loro probabilità linguistica e contestuale;
- Indicazione della probabilità di correttezza delle proposte con una valutazione metrica;
- Visualizzazione fotorealistica delle ricostruzioni testuali nelle lacune sulla base dei caratteri identificati con l'HTR e con highlights dinamici (selezionando una lacuna nel testo si evidenzia la zona dell'immagine corrispondente);
- Feedback utente;
- Possibilità di inserire manualmente integrazioni e confrontarle con quelle AI;
- Possibilità di memorizzare le diverse ipotesi ricostruttive;
- Database per l'archiviazione dei dati.

Anno 3

Semestre 1. Completamento delle funzionalità avanzate di Theion (pagamento V tranche)

La società appaltatrice dovrà fornire versione aggiornata di Theion che includa tutte le funzioni descritte nell'Art. 4, Milestone 3, cioè:

- Fine-tuning delle funzioni sviluppate nella fase Alpha;
- Ottimizzazione dell'algoritmo AI per affinare l'integrazione delle lacune;
- Miglioramento delle spiegazioni sulle alternative: giustificare perché una ricostruzione è preferibile mostrando paralleli testuali e contenutistici individuati tramite text mining;
- Modulo di ricerca paralleli: offrire la possibilità di escludere determinate fonti o filtrarle in base alle preferenze;
- Link diretto a paralleli testuali (da TLG, papyri.info, ecc.);
- Interfaccia più avanzata e personalizzabile;
- Possibilità di esportazione dei dati.
- La società appaltatrice si impegna, altresì, a collaborare con il PI all'organizzazione di un workshop di formazione su Theion con sede presso la Host Institution.

Semestre 2. Rifinitura, testing, e finalizzazione di Theion per il rilascio al pubblico (pagamento VI tranche)

La società appaltatrice dovrà:

- Fornire la versione finale del software stabile e pienamente funzionante, completa di dataset finali, modelli addestrati e codice sorgente, come descritta nell'Art. 4, Milestones 2 e 3.
- Collaborare con il PI alla scrittura di un MOOC (Massive Open Online Course) con istruzioni e video tutorial sull'utilizzo di Theion;

- Partecipare all'evento conclusivo del Progetto.

Al termine di ogni semestre, verificato il completamento delle attività previste per il periodo di riferimento, verrà rilasciata attestazione di regolare esecuzione da parte del DEC e verificata dal RUP. Al rilascio della regolare esecuzione, l'operatore economico potrà emettere regolare fattura per l'importo della tranche semestrale, che verrà liquidata secondo le tempistiche previste per legge.

Si precisa che le attività oggetto del presente contratto dovranno essere ultimate, con consegna di tutti i prodotti previsti e, previa autorizzazione del RUP, l'emissione della relativa fattura finale, entro e non oltre il 20 agosto 2028.

Il mancato completamento delle attività contrattuali e/o il sostenimento delle relative spese oltre tale termine comporterà l'inammissibilità delle stesse ai fini del finanziamento ministeriale a valere sul Fondo Italiano per la Scienza (FIS 2). In tal caso, l'Università sarà esonerata da qualsiasi obbligo di pagamento in relazione a prestazioni, spese o fatture sostenute o emesse oltre il termine sopra indicato.

ART. 5. OBBLIGO DI CONSEGNA DEI DATI NEL CORSO DEL PROGETTO

Al fine di garantire la continuità del progetto e tutelare l'Host Institution in caso di interruzione anticipata dell'attività da parte della compagnia appaltatrice, quest'ultima si impegna a:

- Rendere disponibili periodicamente i dati, i moduli sviluppati e qualsiasi materiale intermedio prodotto durante il corso del progetto. La trasmissione di tali dati dovrà avvenire in corrispondenza di ogni tranche di pagamento, assicurando che l'Host Institution abbia sempre accesso agli sviluppi più recenti.
- Fornire una documentazione dettagliata sullo stato di avanzamento del software, comprensiva di codice sorgente, architettura e risultati intermedi.
- Garantire l'accesso ai repository e alle infrastrutture di sviluppo utilizzate nel progetto.

ART. 6. CLAUSOLA DI RISOLUZIONE E GARANZIE

Nel caso in cui la compagnia appaltatrice non rispetti gli obblighi di consegna progressiva dei materiali e dei dati, l'Host Institution si riserva il diritto di risolvere il contratto e trattenere i materiali consegnati fino a quel momento. Inoltre, in caso di interruzione anticipata del progetto da parte della compagnia appaltatrice, essa dovrà garantire il trasferimento completo di tutti i materiali sviluppati fino a quel momento, senza alcun costo aggiuntivo.

ART. 7. PROPRIETÀ INTELLETTUALE

Tutti i diritti sul software, sui dati e sui materiali sviluppati durante il progetto saranno di esclusiva proprietà dell'Host Institution. La compagnia appaltatrice non potrà in alcun modo utilizzare, distribuire o rivendere il software o i suoi componenti senza l'esplicito consenso scritto dell'Host Institution.

Il Responsabile Unico del Progetto
Dott. Michele Codella