LACUNA – *"Leveraging Innovative Approaches to Comprehensively Understand Ancient Epicurean Texts. Towards the First AI-Enhanced Editions of Herculaneum Theological Papyri"*

Project funded under the Italian Science Fund 2022–2023 (FIS 2 Call).

Project Code: FIS-2023-01833

CUP: E53C24003870001 CUI: S00876220633202500005

TECHNICAL SPECIFICATIONS OF THE TENDER

INDEX

Art. 1 – SUBJECT OF THE CONTRACT

This contract concerns the research activity aimed at the design and development of *Theion*, a software applying artificial intelligence models to the textual reconstruction of Herculaneum papyri, to be carried out over a three-year period by a specialized external company. The ultimate goal of the project is the delivery of a complete and functional software product, compliant with the technical specifications defined by the Host Institution in Art. 3.

Art. 2 – GENERAL CHARACTERISTICS OF THE REQUIRED SOFTWARE

The *Theion* software will be developed within the three-year LACUNA project and will provide an advanced solution for the textual reconstruction of fragmentary Greek papyri from Herculaneum, specifically those containing the theological works of Philodemus of Gadara targeted by the project. As per the work plan described in Art. 3, the software must implement AI models to:

1. Generate linguistic predictions to suggest gap restorations based on context and authorial style. The AI models will be trained on the corpus of ancient Greek texts and, specifically, on the corpus of Philodemus of Gadara, partially available via papyri.info and partially expanded during the project. The models used for predicting missing texts must be capable of using both human input and images of the Herculaneum papyri. Therefore, the software must integrate AI methodologies for automatic character recognition in the images and for segmenting text sequences, with output in the form of diplomatic transcription.
2. Perform automated text mining to suggest relevant textual and content parallels based on linguistic and conceptual similarity criteria.

Art. 3 – WORK PLAN AND PROJECT TIMELINE

The research commission will be carried out through four Milestones distributed over the 3 years of the Project:

MILESTONE 1 (YEAR 1): AI Data Collection; HTR on Papyri Images

Semester 1: AI Training Data Collection

The first semester involves researching and collecting appropriate datasets to train AI models capable of linguistic predictions for analyzing and reconstructing Philodemus' fragmentary Greek papyri. This includes accessing open-source repositories of all available ancient Greek texts. To enable the AI models to recognize Philodemus' distinctive style, syntax, and vocabulary, they must be trained on the complete digital corpus of his known works available on papyri.info, which will be further expanded during the project.

Semester 2: HTR on Papyri Images

Since AI models must incorporate both human input and papyri images, the second semester will focus on integrating AI-based HTR (Handwritten Text Recognition) for multispectral images of selected papyri, including segmenting text sequences and producing diplomatic transcriptions. The possibility of offering a photorealistic reconstruction of gaps in the *Theion* user interface will also begin to be explored. Furthermore, synergies between AI gap-filling models and HTR systems on stratified papyri should be investigated to enhance reconstruction of interrupted text sequences.

MILESTONE 2 (YEAR 2): AI Model Training; *Theion* Development (Alpha Stage)

The second year of the project will be dedicated to two activities, organized over two semesters:

Semester 1. Training of AI models for gap integration and for text mining of lexical and content-based parallels.

- Gap integration: To assess the predictive accuracy of the models, the most suitable approaches must be evaluated, including: comparison between texts with actual gaps and complete texts (creating fake gaps in complete texts to test the model's ability to predict the missing segments); comparison between AI-proposed integrations and human-made integrations. Interaction between the contractor's team and the papyrological team of the Project is crucial at this stage.
- Text mining of parallels: The AI models will apply automated text mining to suggest relevant textual and conceptual parallels that can assist scholars in the editorial process. The identification of parallels should not rely on word-for-word matching (as already possible using the "Text search" and "Wildcard search" functions of the TLG), but rather on linguistic, stylistic, and conceptual similarity criteria.

Semester 2: *Theion* Alpha Development.

Creation of the prototype with core functionalities, focused on gap reconstruction and the management of alternatives. At this stage, *Theion* should:

- Have a basic software infrastructure that allows for the uploading of texts and images;
- Offer an integrated visualization in which the following are displayed side by side: the multispectral image of the papyrus, the transcription (entered by the scholar and/or generated by the HTR), and the gap-filled text proposed by the AI;
- Perform automatic prediction of gap fillings based on the available space in the gap, the Greek language, the passage's context, the author's style, and the identified parallels;
- Display the alternatives ranked according to their linguistic and contextual probability;
- Indicate the likelihood of correctness of the suggestions with a metric evaluation;
- Provide a photorealistic visualization of the textual reconstructions in the gaps based on the characters identified via HTR, with dynamic highlights (selecting a gap in the text highlights the corresponding area of the image, and vice versa);
- Allow users to report errors or improve the suggestions through feedback;
- Enable manual input of reconstructions and comparison with AI-generated ones;
- Store different reconstruction hypotheses;
- Include a database for data storage.

MILESTONE 3 (YEAR 3): *Theion* Beta Development
Semester 1: Completion of Advanced Features

- Fine-tuning Alpha features;
- Optimizing AI for improved gap filling;
- Enhancing justification of reconstruction alternatives through mined parallels;
- Adding filters/exclusion options for sources in parallel search;
- Providing direct links to parallels (TLG, papyri.info, etc.);
- Offering a customizable interface;
- Enabling data export.

Semester 2: Refinement, Testing, and Final Release

- Refining the user interface;
- Integrating user feedback.

MILESTONE 4 (YEARS 1–3): Dissemination of Results

- Years 1–3: monthly meetings between the contractor's team, the PI, and the Project's papyrological team;
- Years 1–3: submission of articles to journals in the fields of papyrology, computer science, and/or Digital Humanities;
- End of Year 2 / Beginning of Year 3: training workshop on Theion;
- By the end of the project: development of a MOOC (Massive Open Online Course) with instructions and video tutorials on how to use Theion;
- Participation in the Project's final event.

ART. 4. SERVICE DELIVERY AND PAYMENT TERMS

The awarded company must ensure the progressive development of the software through periodic releases. Payment of the awarded amount, relative to the base tender amount (€300,000 excluding VAT), will be divided into six semi-annual tranches, each equal to 1/6 of the awarded amount, which will be disbursed upon completion of the activities described in Art. 3 under the following conditions:

Year 1
Semester 1. Data collection for AI training (payment of 1st tranche)
The contractor must provide a complete and detailed list of datasets collected for training the AI models.
Semester 2. HTR on papyrus images (payment of 2nd tranche)
The contractor must provide:

- Functional HTR software applied to the multispectral images of the selected papyri;
- Sample images annotated with diplomatic transcriptions obtained via the HTR system.

Year 2
Semester 1. Training of AI models for gap integration and text mining (payment of 3rd tranche)
The contractor must provide:

- AI models trained for gap integration and text mining on the complete digital corpus of known works by Philodemus;
- Evidence of testing on real gaps and "fake gaps";
- Comparison between AI outputs and human integrations (including documentation of methods and results);
- Documented results of automated text mining with examples of identified parallels.

Semester 2. Development of Theion – Alpha phase (payment of 4th tranche)

The contractor must deliver a working prototype of Theion including all features described in Art. 3, Milestone 2, Semester 2, namely:

- Core software infrastructure for uploading texts and images;
- Integrated display aligning: multispectral image of the papyrus, the transcription (input by scholar and/or generated by HTR), and the AI-proposed integrated text;
- Automatic gap integration predictions based on gap space, Greek language, textual context, author style, and identified parallels;
- Visualization of alternative reconstructions ranked by linguistic and contextual probability;
- Indication of confidence level for each proposal, with metric evaluation;
- Photorealistic display of textual reconstructions within gaps using characters identified via HTR, with dynamic highlights (selecting a gap in the text highlights the corresponding image area);
- User feedback functionality;
- Manual input of integrations and comparison with AI outputs;
- Ability to store different reconstruction hypotheses;
- Database for data storage.

Year 3

Semester 1. Completion of Theion's advanced features (payment of 5th tranche)

The contractor must deliver an updated version of Theion including all features described in Art. 4, Milestone 3, namely:

- Fine-tuning of features developed during the Alpha phase;
- Optimization of the AI algorithm to improve gap integration;
- Enhanced explanations of alternatives: justifying preferred reconstructions by displaying identified textual and conceptual parallels through text mining;
- Parallels search module: enabling exclusion or filtering of sources based on user preferences;
- Direct links to textual parallels (e.g., from TLG, papyri.info, etc.);
- More advanced and customizable interface;
- Data export capability.
- The contractor will also cooperate with the PI to organize a training workshop on Theion at the Host Institution.

Semester 2. Finalization, testing, and public release of Theion (payment of 6th tranche)

The contractor must:

- Deliver the final stable and fully functioning version of the software, complete with final datasets, trained models, and source code, as described in Art. 4, Milestones 2 and 3;
- Collaborate with the PI in drafting a MOOC (Massive Open Online Course) with user instructions and video tutorials for Theion;
- Participate in the final event of the Project.

At the end of each semester, upon verification of the completion of scheduled activities, a certificate of regular execution will be issued by the DEC and validated by the RUP. Upon issuance of this certificate, the contractor may issue an invoice for the relevant tranche, which will be paid according to the timelines set by law.

It is specified that the activities covered by this contract must be completed, with delivery of all expected products and, subject to RUP's authorization, issuance of the related final invoice, no later than **August 20, 2028**.

Failure to complete the contractual activities and/or incur related expenses by this deadline will result in their ineligibility for ministerial funding under the Italian Science Fund (FIS 2). In such cases, the

University will be exempt from any obligation to pay for services, expenses, or invoices incurred or issued after the aforementioned deadline.

Art. 5 – OBLIGATION TO DELIVER DATA DURING THE PROJECT

To ensure project continuity and protect the Host Institution in case of early termination, the contractor agrees to:

- Periodically provide all data, modules, and intermediate materials produced. These must be transmitted with each payment tranche, ensuring the Host Institution has access to the latest developments;
- Provide detailed documentation on software progress, including source code, architecture, and intermediate results;
- Guarantee access to project repositories and development infrastructure.

Art. 6 – TERMINATION CLAUSE AND GUARANTEES

If the contractor fails to meet the obligation of progressive delivery of materials and data, the Host Institution reserves the right to terminate the contract and retain the delivered materials. In case of early termination by the contractor, they must ensure full transfer of all developed materials at no extra cost.

Art. 7 – INTELLECTUAL PROPERTY

All rights to the software, data, and materials developed during the project will be the exclusive property of the Host Institution. The contractor may not use, distribute, or sell the software or its components without the Host Institution's explicit written consent.

Project Manager
Dr. Michele Codella