# Label the Invisible: AI-Aided Label Enhancement and Ink Residue Exposure

Nadiia Duiunova[1,*,†], Mariia Halchynska[1,†], Johannes Römisch[1,†], Hannes Kahl[2], Holger Essler[3] and Frank Deinzer[1]

[1]Technical University of Applied Sciences Wuerzburg-Schweinfurt, Center for Artificial Intelligence (CAIRO), Franz-Horn-Straße 2, 97082 Würzburg, Germany

[2]University Trier, Department III, 54286 Trier, Germany

[3]Julius Maximilian University of Würzburg, Institute of Classical Philology, Chair of Classical Philology I (Greek Studies), Residenzplatz 2, 97070 Würzburg Germany

**Abstract**

The carbonized Herculaneum scrolls represent a unique challenge for text recovery due to their fragile state and the visual similarity between ink and papyrus substrate. This study presents an iterative, human in-the-loop approach for ink detection in Scroll 5 (PHerc. 172) using a TimeSformer-based deep learning model. Building upon preprocessed X-ray Phase-Contrast Tomography (XPCT) data from the Vesuvius Challenge, we employ progressive model refinement and expert-verified labeling to improve the identification of ancient Greek letterforms. Our approach demonstrates that precise, trace-based annotation combined with repeated training cycles leads to a significant increase in legible character recognition from 169 to 368 letters across 15 scroll segments. The results underscore the value of interdisciplinary collaboration and iterative feedback in advancing the digital decipherment of ancient texts.

**Keywords**

X-ray phase-contrast tomography, transformer-based ink detection, iterative model refinement, deep learning for low-signal data, ancient document analysis,

## 1. Introduction

The eruption of Mount Vesuvius in 79 AD sealed away a unique literary treasure: the carbonized papyrus scrolls from the Villa of the Papyri in Herculaneum. Preserved under layers of volcanic material, these ancient manuscripts offer a rare glimpse into the intellectual life of the Roman world. Yet for centuries, their charred and fragile state rendered many of them unreadable. Recent advancements in non-destructive imaging, particularly X-ray Phase-Contrast Tomography (XPCT), have created a new opportunity to digitally access their contents without physical unrolling.

To advance this opportunity, the Vesuvius Challenge [1] was launched in 2023 as an open competition aimed at deciphering the hidden texts of the Herculaneum scrolls and to accelerate this endeavor by providing researchers with high-resolution XPCT data and the tasks of geometric reconstruction by virtual unwrapping of the scrolls and detecting and reconstructing ink traces hidden within the scrolls. The challenge has stimulated the development of machine learning pipelines capable of enhancing text visibility and separating faint ink from the dense papyrus background. Among the most significant breakthroughs has been the adaptation of TimeSformer, a transformer-based vision model [2], to the domain of volumetric ink detection. Originally trained on labeled segments from Scroll 1 (PHerc.Paris. 4), this model serves as a foundational tool in our study.

*Perspectives on Humanities-Centred AI and Formal & Cognitive Reasoning Workshop 2025,(CHAI 2025 & FCR 2025), Joint Workshop at the 48th German Conference on Artificial Intelligence, September 16, 2025, Potsdam, Germany*

*Corresponding author.

†These authors contributed equally.

nadiia.duiunova@study.thws.de (N. Duiunova); mariia.halchynska@study.thws.de (M. Halchynska); johannes.roemisch@study.thws.de (J. Römisch); kahlh@uni-trier.de (H. Kahl); holger.essler@uni-wuerzburg.de (H. Essler); frank.deinzer@thws.de (F. Deinzer)

0009-0009-7483-2567 (N. Duiunova); 0009-0008-8380-0172 (M. Halchynska); 0009-0002-8037-8857 (J. Römisch); 0009-0002-2634-9967 (H. Kahl); 0000-0003-4759-7716 (H. Essler); 0000-0003-2573-3529 (F. Deinzer)

In this study, our primary focus is on Scroll 5 (PHerc. 172), where we utilize the preprocessed outputs of the Vesuvius Challenge to train and refine an ink detection system based on the TimeSformer model. The proposed methodology involves a human-in-the-loop approach, wherein initial model predictions are systematically reviewed and refined by domain experts. These refinements then serve as a foundation for subsequent training iterations. This iterative feedback loop has been demonstrated to result in progressive improvements in model accuracy and output clarity.

The primary metric employed is the quantity of clearly identifiable Greek letters, as verified by a specialist, across successive model iterations. This approach maintains a balance between technical precision and historical authenticity, ensuring that detected ink traces correspond to plausible ancient letterforms. The results demonstrate that through careful annotation and iterative refinement, even faint textual signals embedded within the scrolls can be successfully recovered and rendered legible.

The remainder of this paper is structured as follows: Section 2 provides an overview of related work, including past Vesuvius Challenge contributions and the historical development of letter shape analysis and transformer-based models. Section 3 introduces the dataset and preprocessing pipeline, with a focus on XPCT scanning and the segmentation method used to flatten the scroll layers. Section 4 details our methodological approach, including labeling strategies, iterative training, and implementation details. In Section 5, we present our experimental results and evaluate model performance across iterations. Section 6 concludes the paper with a discussion of findings and outlines directions for future research, including semi-supervised techniques and improved preprocessing workflows.

## 2. Related Work

The data of the Vesuvius Challange is an application of computer tomographic (CT) measurement and imaging in the context of the evaluation of man-made artifacts. The description and estimation of these artifacts is related to man-made scientific reading and assessment schemes. In its entirety it covers topics of material distinction and description, physical, chemical description of material change under conditions of temperature and time, material representation through measurement and its computational estimation, also the description of change of human labor habits related to production and usage of papyrus, the development of writing, reading and storing information. Out of that we choose three aspects, that we will cover in the section of related work: (1) Ink on charred papyrus, (2) script usage and (3) the computation on the data (model and modelling) in relation to ink residues.

### 2.1. Ink on charred papyrus

In 2015 the founders of the challenge published their research on a charred scroll (parchment) from En-Gedi [3], they showed significant progress in evaluation of ink residues in the XPCT datasets. They express that the ink composition must have involved iron or lead. Which is not the whole chemical truth, because it is iron on the atomic level, but a metal oxide on the molecular level (in respect to the burning [more important since the ferric gallat is solid up to 170 °C and turns to iron oxide at 350 °C [3]] and the time of being buried). The analysis pipeline built on previous work on the Herculaneum papyri from 2009 and 2013. In the year 2015 there was also research published on the ink of the papyri from Herculaneum [4], [5] and tracking ink composition on Herculaneum papyrus scrolls [6]. As we learn from these perspectives it is likely that not only metallic inks were used to write the scrolls. The research offered the obvious insight, that the CT method is not the most suitable here, an infrared photography would suit the problem better (reflection vs. absorption measurements uncover surface structures [even if they are made up of equal material]), but unfortunately it is not available for non destructive workflows. As a result of that research period the presents of metal in ink was taken for granted, in contrast to the archaeologists' assumption of purely carbon based ink, for the Herculanum papyri. In the end it is not clear where the metal components of the ink come from and a biotic source of plant material (the main component of carbon ink) is not improbable (find PCA Cluster of charcoal from different specimen [7]; a general overview to the scientific outcome about ash and charcoal [8], on the other hand the Pb content was discussed in terms of polluted water, but not in terms of accumulation in biomass and thus high Pb content in biomass ash and charcoal [9]).

### 2.2. Script usage

For the purpose of this paper it is essential to know potential form and development of letters used in the Herculaneum scrolls. They span from the fourth century BCE to 79 CE with Latin and Greek letters. A detailed treatment is given by Cavallo, "Libri scritture scribi a Ercolano", Naples 1983 [10]. In our case the letters are Greek majuscules written by a literary hand.

### 2.3. TimeSformer ink model

The original analysis pipeline (mesh model of scroll as a mass spring system driven by the gradient criterion of second order symmetric tensor and salience measure of the intensity volume (which denotes the location of animal skin surface, mapping voxels as textures to mesh model, flatten segments, merge segments by hand)) is now dominated by AI-driven methods. The reason for that is threefold: first the segmentation of sheets of material is difficult and local geometrical and statistical expressions of surface orientation are not always possible depending on the preservation state (some papyrus sheets are simply crushed or baked together); second the poor ink signal in the XPCT dataset needs to be boosted and third it's the AI time. We will focus on the work that was carried out to estimate ink residues in flattened parts of a papyrus scroll (segments) represented by the volume around of the papyrus sheet. To solve the problem of estimating regions of an unknown material density characteristics in relation to surrounding regions and classify them as written regions. Youssef Nader utilizes the training of the a TimeSformer model and its application for inference. The approach is quite different from the original intention of the TimeSformer [11]. The first difference is, that any autoencoder of the input data is neglected. The data is used directly with some additional masking. Second difference is, that the decoder stage of the timeSfomer is not used. That means the encoder reproduces masks for a given intensity volume. That is fine, because due to the fact, that ink was nearly invisible on the intensity level, the only human driven labeling is possible in rare situations (crackles). And labeling on top of the encoder output was the only solution to coupe with small intensity gradients representing papyrus and ink in contrast to papyrus only.

## 3. Dataset and Preprocessing

The analysis of ancient carbonized scrolls, particularly those recovered from the Villa of the Papyri in Herculaneum, requires a digital pipeline to prepare the data for ink detection and text reconstruction. We leverage the preprocessed volumetric data provided by the Vesuvius Challenge [1], which includes high-resolution scans, segmentation outputs, and flattened papyrus sheets. Understanding the preceding preprocessing steps is crucial for contextualizing our methodology and its constraints. This section outlines the key stages in data acquisition and preparation, highlighting the scan process and the segmentation methodology provided through the Volume Cartographer [12] and Thaumato Anakalyptor [13].

### 3.1. Scan Process

The scrolls examined by the Vesuvius Challenge community have undergone a non-destructive scanning procedure using XPCT, executed at the Diamond Light Source synchrotron facility. Unlike conventional absorption-based X-ray imaging, XPCT is sensitive to minute changes in the refractive index, making it ideal for detecting the subtle structural differences within the carbonized papyrus and ink [14, 15]. The result is a three-dimensional volumetric dataset capturing the internal microstructure of the scrolls at a resolution sufficient to resolve the fine layers (as seen on Figure 1) and potential ink traces embedded within.

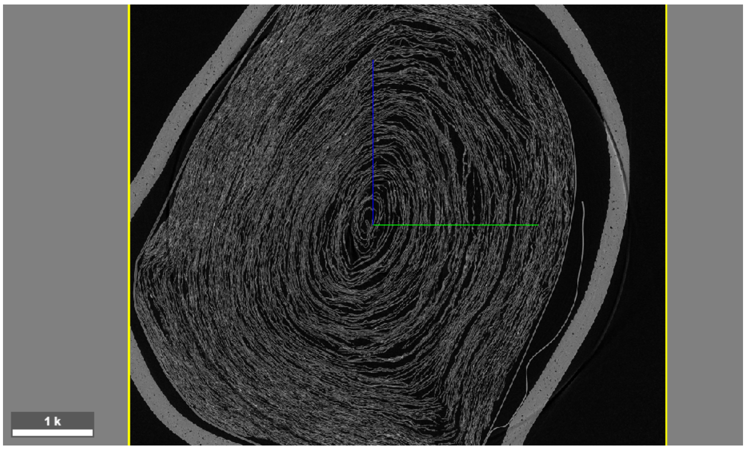


**Figure 1:** XPCT scan of Scroll 5, showing its internal layered structure.

These raw scans, however, do not directly yield readable text. Due to the compressed and irregular geometry of the scrolls – folded, coiled, and damaged over centuries – text is often embedded within highly distorted 3D spaces, distributed across dozens of sub-millimeter depth layers. Thus, a critical step involves virtual unwrapping, in which the topology of the scroll is modeled computationally and then digitally flattened into a series of depth images suitable for further analysis. This unwrapping pipeline builds on the prior work by Stabile et al. [14] and Baum et al. [16], and has been refined and standardized as part of the infrastructure of the Vesuvius Challenge.

The scroll is ultimately converted into a stack of 65 axial slices – each representing a distinct depth layer within the material [17, 1]. This layered representation, while still volumetric in nature, allows traditional 2D and 3D machine learning tools to be employed for segmentation and ink prediction.

### 3.2. Segmentation and Flattening

One of the main components of the Vesuvius Challenge infrastructure is the Volume Cartographer pipeline, introduced by Parsons et al. [18], which automates the segmentation and mapping of scroll geometries onto 2D image stacks. In contrast to earlier manual or semi-automated methods, Volume Cartographer provides a reproducible, data-driven solution to the segmentation problem, producing aligned sheet maps from raw volumetric data.

At the core of Volume Cartographer is a hierarchical sheet detection algorithm that identifies and tracks distinct papyrus layers within the scroll volume. The system begins by isolating regions of coherent planar structure using a curvature-based filter across the volumetric scan. These candidate layers are then refined through a multi-stage fitting process that leverages surface normals and local geometry to produce high-fidelity surface meshes. These meshes are subsequently used to interpolate flattened representations of each sheet via a deformation-minimizing transformation. The output is a stack of 2D orthogonal slices that correspond to the original layers but have been geometrically corrected to remove bending and torsion [18].

In our project, we build upon the pre-segmented and virtually unwrapped output generated by Volume Cartographer. This includes the depth-normalized 65-slice representation of Scroll 5, from which we extract regions of interest for ink detection. Importantly, the segmentation process already incorporates denoising and artifact rejection, meaning that the layers we process have reduced geometric noise and exhibit a relatively consistent topology [1, 18].

While our contribution does not involve segmentation itself, the quality and accuracy of the input data are directly contingent on the success of Volume Cartographer. Misalignments, topological artifacts, or segmentation errors at this stage can propagate downstream, leading to false positives in ink detection, i.e. the misidentification of substrate textures as letterforms. Therefore, the robustness and generality of Volume Cartographer were essential prerequisites for the iterative modeling pipeline described in subsequent sections.

Moreover, the segmentation approach offers an implicit form of contextual regularization. By aligning layer surfaces according to the physical curvature of the scroll, Volume Cartographer ensures that adjacent image slices correspond to adjacent material layers, preserving local spatial continuity. This is particularly valuable for models like TimeSformer [2], which rely on consistent spatiotemporal correlations to detect faint and dispersed ink traces across slices. These processed datasets form the basis of our study, allowing us to bypass the computationally intensive segmentation phase and focus directly on ink inference.

By building atop the outputs of XPCT scanning and the Volume Cartographer system, we are able to engage directly with virtual representations of ancient manuscripts, thus sidestepping the most fragile and destructive aspects of physical papyrology. These processed scroll images provide a reliable substrate upon which our ink inference techniques are deployed.

# 4. Methodology

Our methodology aims to reliably detect ink traces from the segmented, flattened volume slices produced during preprocessing. Given the extreme visual similarity between carbon-based ink and the papyrus substrate, we adopt a human-in-the-loop strategy that combines expert-informed manual labeling, iterative model refinement and implementation techniques designed to maximize visibility and interpretability. This section illustrates the system's implementation details, our image enhancement and labeling strategies and the iterative model training workflow.

## 4.1. Implementation Details

All models were implemented using the PyTorch Lightning framework to enable modular and reproducible training. We used Weights & Biases [19] to track all experiments, logging metrics, hyperparameters, and qualitative outputs.

Training was conducted on a remote Linux server equipped with an NVIDIA Tesla V100 GPU (32GB VRAM). We adopted the 2023 grand prize-winning method from the Vesuvius Challenge, which utilizes a TimeSformer model architecture [17] using a sliding window approach. Each training input consisted of a $64 \times 64 \times 26$ volumetric window, extracted with a stride of 32. The very small scale of this window size is visualized in Figure 2 as a yellow 64 by 64 pixel square, which takes up only a fraction of the space of the full letter. This configuration is chosen to preserve local structural context of the papyrus material while preventing the model from being able to learn to recognize entire characters, which would likely lead to overfitting and poor generalization.

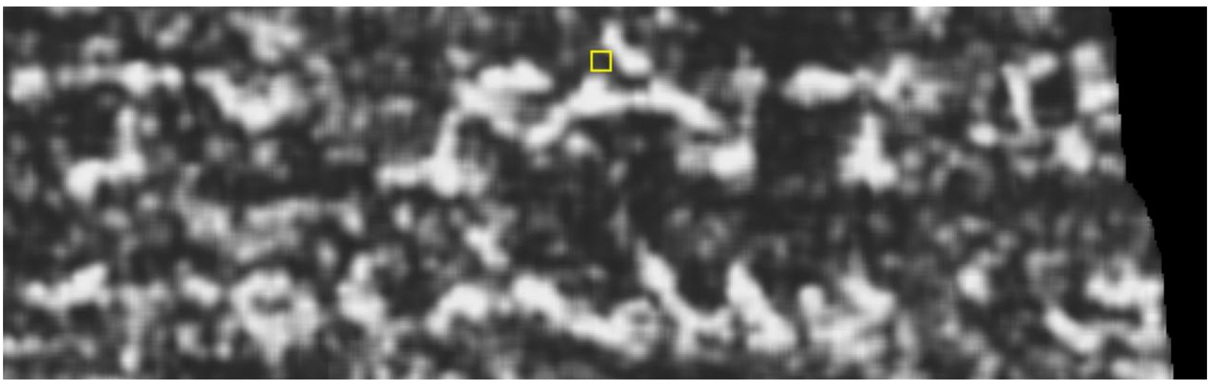

**Figure 2:** Visualization of window size (yellow square) on inference result with initial model on segment 20241030083650

To evaluate the model performance during training, the segment 20241028121111 (see Figure 3) was always held out as a validation set and never used for training. This segment overlapped substantially with a nearby labeled segment (20241028121112), minimizing overall data loss while still providing the trainer class with metrics for evaluation and learning rate scheduling.

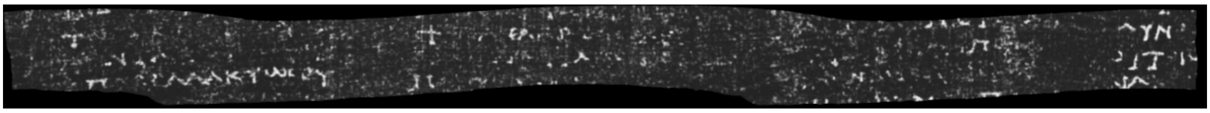

**Figure 3:** Model output on held-out validation segment 20241028121111 after final training iteration

## 4.2. Labeling Techniques

To enhance the interpretability of scroll segments generated during virtual unwrapping, we began with predictions from a pretrained TimeSformer model [17]. These initial predictions served as a starting point for expert-guided manual annotation.

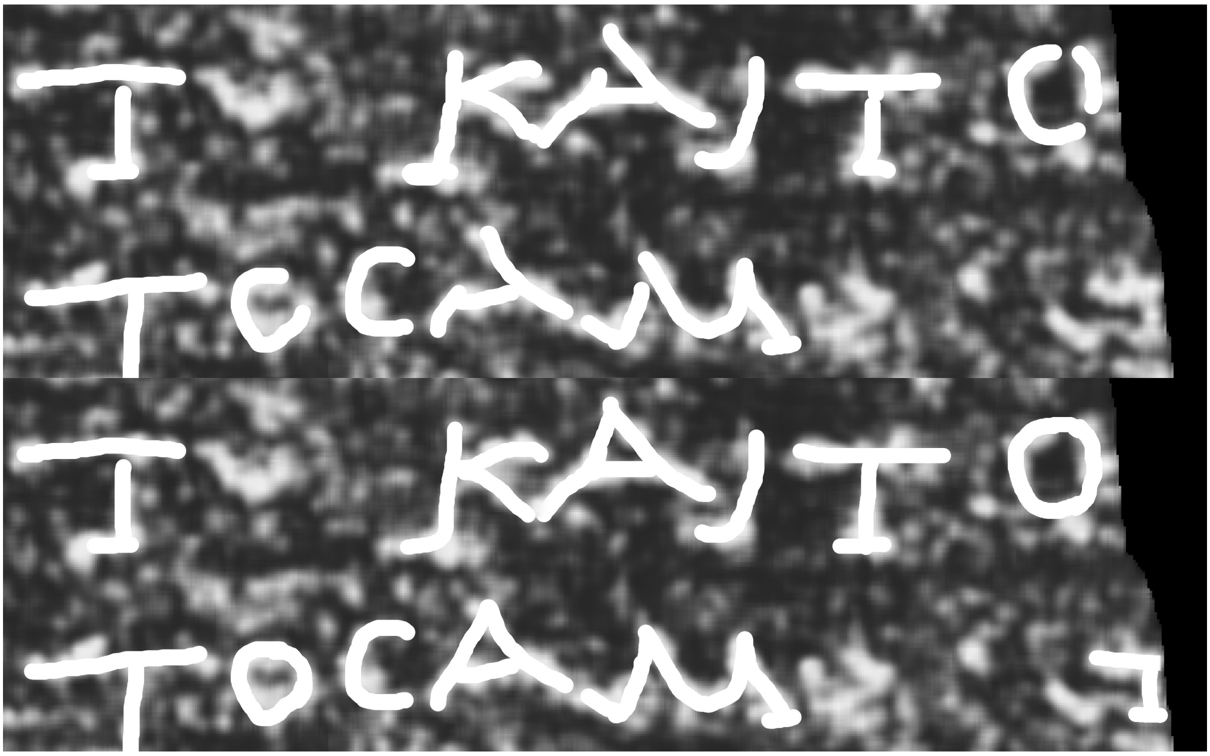

**Figure 4:** Comparison of two labeling strategies Trace-based (top) and Expansive (bottom) on Inference result with initial model on segment 20241030083650

To create a reliable training set, we considered two alternative labeling strategies:

- **Trace-based labeling:** This conservative strategy prioritized precision. We labeled only clearly visible, well-defined ink traces that matched expected Greek letterforms. Faint traces were included only if they structurally aligned with expected shapes and showed sufficient contrast against the background. All labels were validated by a specialist of ancient Greek, referencing authoritative sources including [20] and comparative paleographic studies [21].
- **Expansive labeling:** This strategy sought to increase coverage by extrapolating partial traces into full letterforms when there was high confidence in the inferred structure. Ambiguous features – such as vertical strokes potentially being part of multiple characters like $\iota$ (iota), $\Pi$ (pi), $\Psi$ (psi), $\tau$ (tau), $\kappa$ (kappa) – were excluded. This approach increased the quantity of labeled data, as gaps in incomplete letters were filled, as seen in Figure 4. Conversely, it also introduced the possibility of adding false positive labels, which are correct from a reader's perspective, but might teach the model to hallucinate ink, where there is none, ultimately preventing our goal of visual clarity.

Testing the second approach for one iteration, the results were visually much noisier, which led to our choice of labeling using the Trace-based approach for the remainder of the process.

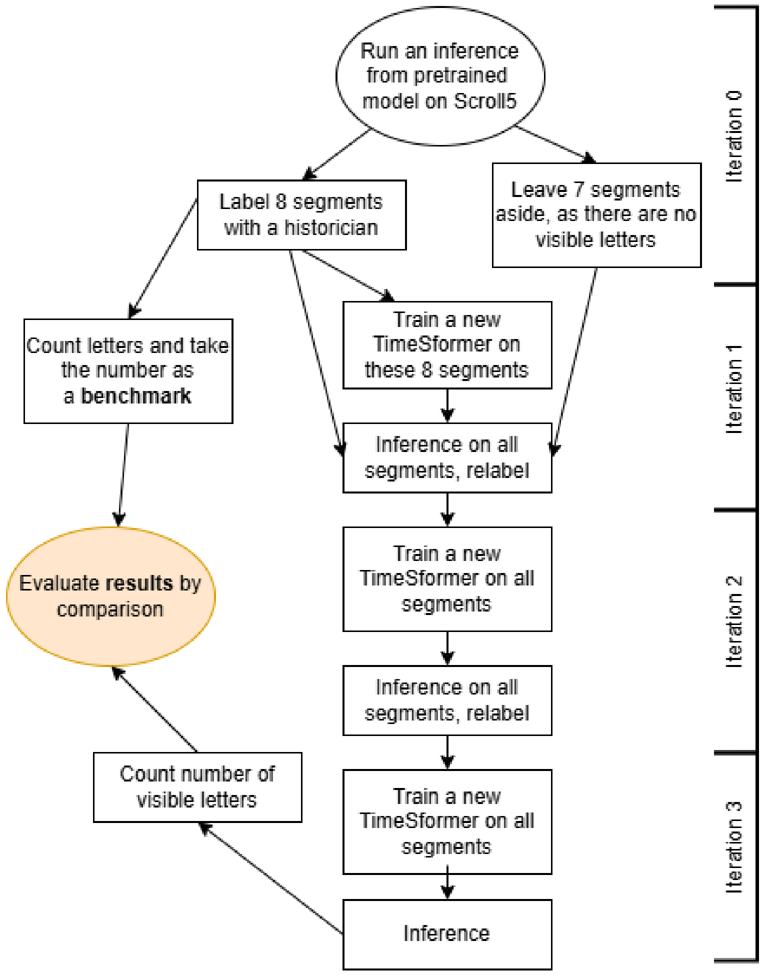


**Figure 5:** Workflow overview of the labeling and iterative model training process.

## 4.3. Progressive Visibility Enhancement of Letter Structures

We employed an iterative model training approach to improve ink trace detection over time. Figure 5 illustrates the overall flow. The central idea was to refine the labels after each iteration based on model inference and expert analysis, thereby enhancing both label quality and model accuracy. Each iteration followed this general process:

1. Correct false positives by removing previously labeled letters that were no longer supported by model inference in the current iteration.
2. Identify and annotate new traces that plausibly match letterforms found in the [20], particularly those attributed to works by Philodemus ($1^{st}$ century B.C).
3. Count the number of readable letters in each segment to evaluate progress.
4. Retrain the TimeSformer [17] model from scratch on the revised label set, then run inference to initialize the next round.

This process allowed us to iteratively improve both the precision of the model and the clarity of the training data. The combination of structured annotation, expert validation, and feedback-guided training significantly reduced noise and false positives.

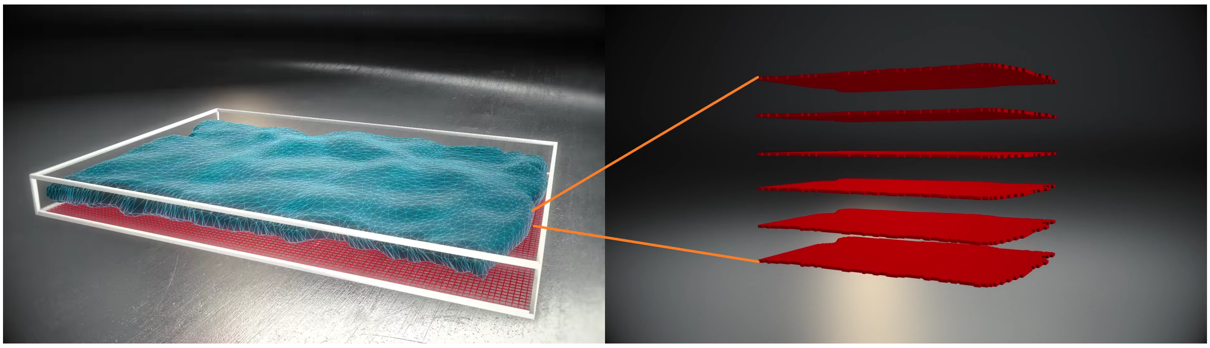

**Figure 6:** CT Scan of papyrus decomposed to 65 Layers. (https://scrollprize.org/tutorial3).

The iterative process concluded after four rounds, when the final iteration no longer showed noticeable improvements in visual clarity or letter count. This saturation point was considered the effective limit of our current approach.

# 5. Experimental Results and Evaluation

## 5.1. Data setup

The data used in this study originates from Scroll 5 of the Vesuvius Challenge dataset [18]. The scroll has been CT-scanned and virtually unwrapped using the state-of-the-art pipeline provided by the competition organizers (explained in Section 3.1).

Although the scroll contains 23 available segments, we worked with a subset of 15, which were used consistently throughout all iterations of the project. These 15 segments include: 20241028121111, 20241028121112, 20241030083650, 20241030152031, 20241105112301, 20241113070770, 20241113080880, 20241113090990, 20241025062040, 20241025145341, 20241025150211, 20241102160330, 20241108111522, 20241108120731, and 202411081207312. Of these, the initial round of manual labeling and model training used the first 8 segments (see Figure 5 for the workflow diagram).

Each segment contains 65 volumetric layers, corresponding to depth-wise slices through a single sheet of papyrus. We consistently used layers 17 through 42, which cover the central portion of each sheet. This range provided the highest ink visibility while avoiding distortion or noise from overlapping neighboring layers. Figure 6 illustrates how one sheet of papyrus appears as multiple layers in the CT data.

As a starting point, we applied the 2023 Vesuvius Challenge grand prize model [17], trained on Scroll 1, to all 15 segments of Scroll 5. The resulting predictions served as input for the first round of manual labeling and model training in our iterative pipeline.

## 5.2. Results

We evaluated the progression of identifiable letters across three model iterations, using a benchmark of 169 letters manually labeled across 8 scroll segments. These letters were verified by an expert in ancient Greek and served as the baseline for subsequent inference and retraining cycles.

Only ink traces that could be confidently identified as full letters of the ancient Greek alphabet were included in the count. Ambiguous fragments – such as vertical strokes that could plausibly belong to multiple characters, like $\iota$ (iota), $\Pi$ (pi), $\Psi$ (psi), $\tau$ (tau), $\kappa$ (kappa) – were deliberately excluded to maintain a high standard of certainty in the evaluation.

In the first iteration, inference was performed on all 15 segments, resulting in 286 identifiable letters – an increase of 117 over the initial benchmark. This improvement was due to the model's ability to generalize beyond the initial labeled set and detect additional partial or faint characters. Iteration 2

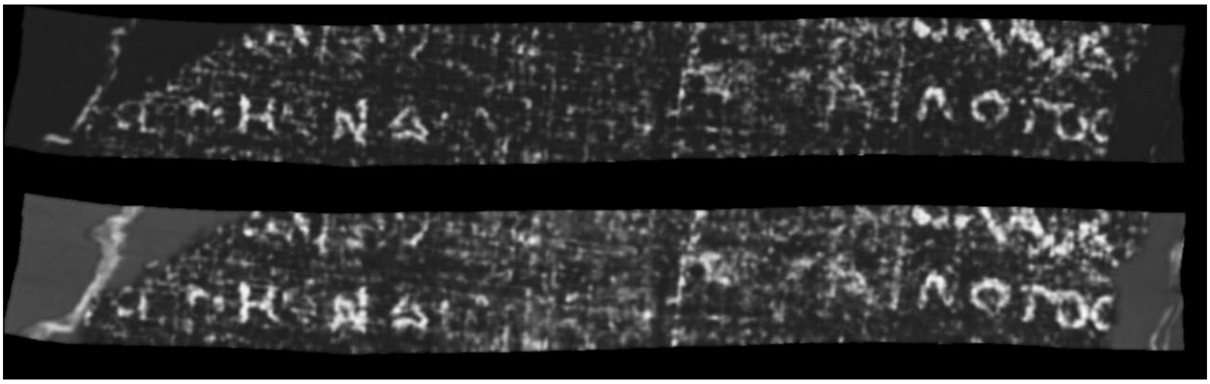

**Figure 7:** Transformation of letters in segment 20241113070770 from ΛΟΠΧ to ΛΟΤΟϹ

yielded a more modest increase to 321 letters, while the final iteration raised the count to 368, marking the highest number of identified characters achieved during the project. Overall, this represents more than a twofold increase compared to the initial benchmark.

While gains diminished across iterations, Iteration 3 still contributed 47 newly recognized letters, indicating that the model continued to extract signal from previously ambiguous areas. However, it also became evident that the majority of readable content had been recovered: noise was largely suppressed, and regions previously confused for ink had been ruled out. This suggests that further iterations would likely not yield meaningful new information from the existing input data.

Performance varied across segments. Some, such as 20241025062040 and 20241108111522, initially had no identifiable letters but saw substantial improvement after refinement (gaining 42 and 22 new letters, respectively). In contrast, other segments with clearer early signals, like 20241030152031, showed slower but steady increases over time. These differences reflect the physical variability in preservation (and the state of the estimation of the target volume of Volume Cartographer) and legibility across the scroll.

While the overall trend across iterations was positive, we did observe a few instances of regression, where the number of identifiable letters in certain segments temporarily decreased. These cases typically arose when regions initially labeled as letters were later judged, by the model or during manual verification, to be more likely noise. As the model improved, it became more selective, sometimes discarding partial or ambiguous traces that were previously accepted as valid characters. In other cases, slight changes in predicted shapes rendered a previously plausible letter unrecognizable. Such corrections, although reflected as numerical decreases, often represented a gain in labeling precision.

Additionally, across iterations, we observed that some characters underwent shape reinterpretation, which contributed to fluctuations in the number of identifiable letters. As shown in Figure 7, an example of such a transformation involved ink traces that were initially classified as Π (Pi) and X (Chi), which in this specific case were later reassigned as T (Tau), O (Omicron), and Σ (Sigma), the latter of which was commonly written in a form resembling the modern Latin letter 'C' during the relevant historical period. While such transformations affected local letter counts within specific segments, they also highlight the model's ability to refine its interpretation of ambiguous traces over time.

Figure 8 and Figure 9 illustrate the evolution of segments 20241108111522 and 20241113080880 respectively, which progressed from zero recognized letters to 22 by the third iteration and from 4 to 14. A full breakdown of results by segment and iteration is provided in Table 1.

While the number of identifiable letters is not a standardized metric, it offers a practical proxy for model performance in this interdisciplinary context. Unlike traditional benchmarks, the goal here is not classification accuracy but the recovery of legible ancient text. These results highlight the effectiveness of iterative refinement for enhancing model reliability in low-signal, high-noise settings, such as carbonized papyrus scrolls.

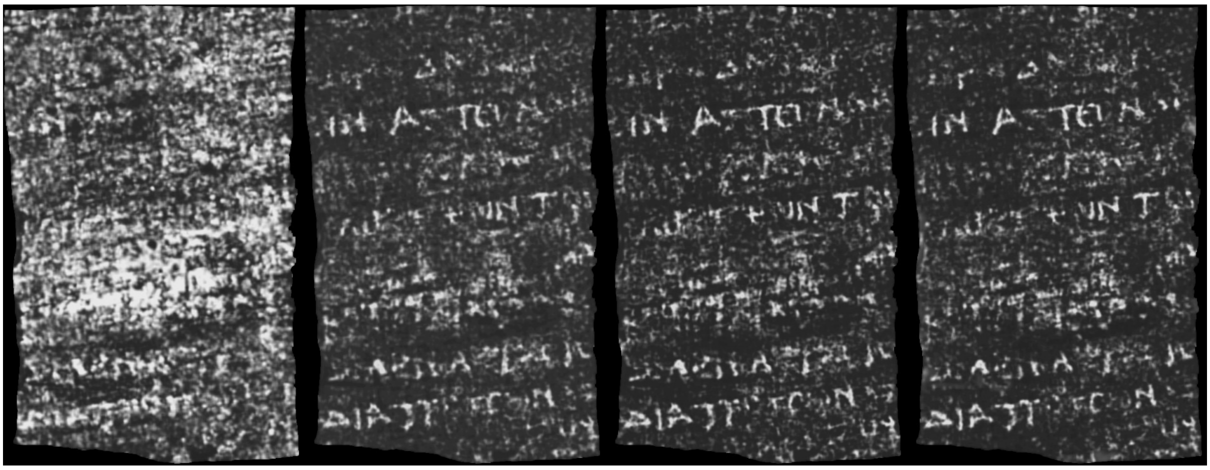

**Figure 8:** Progress of segment 20241108111522 from iteration 0 (left) to iteration 3 (right)

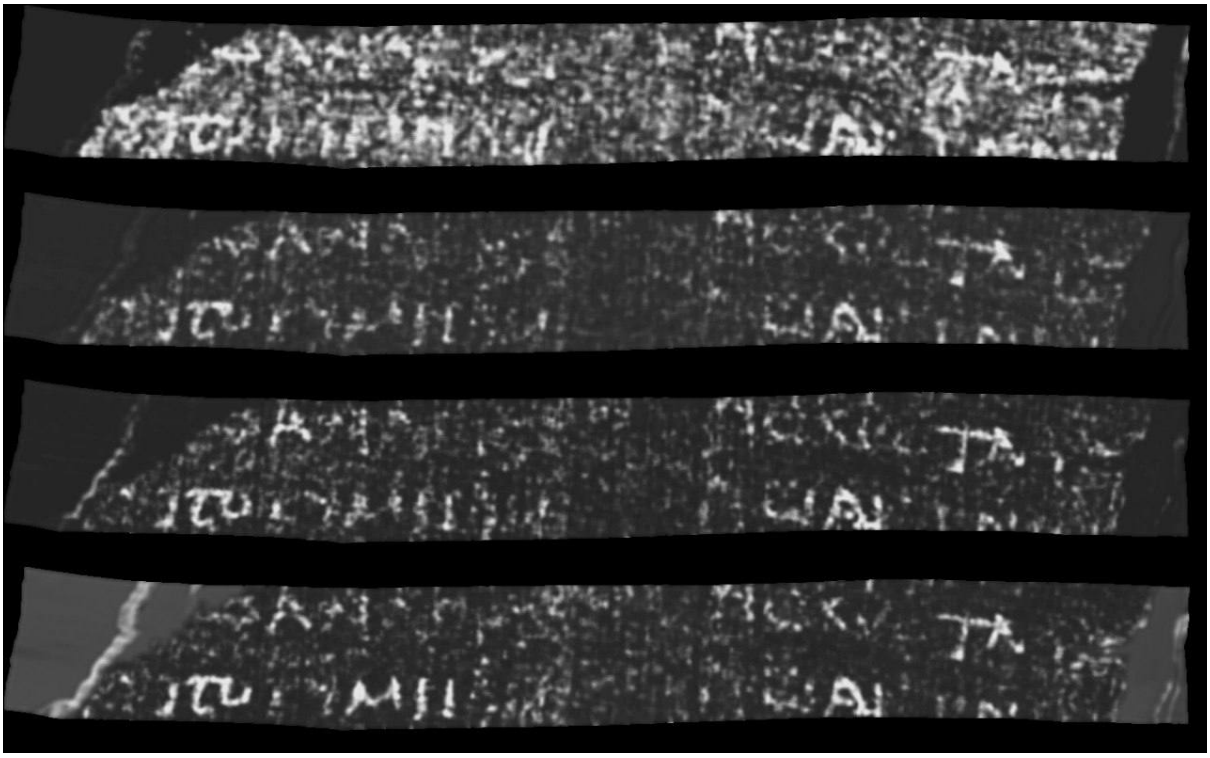

**Figure 9:** Progress of segment 20241113080880 from iteration 0 (top) to iteration 3 (bottom)

## 6. Conclusion and Future Work

This study explored an iterative labeling approach for ink detection in the carbonized Herculaneum Scroll 5. By leveraging a pretrained TimeSformer model, originally trained on Scroll 1, we systematically refined our dataset through multiple cycles of manual annotation and model retraining. This process resulted in a significant increase in the number of identifiable letters, demonstrating the effectiveness of progressive label refinement in challenging historical document analysis tasks.

Our findings highlight the importance of high-quality manual annotation in improving model performance, particularly in cases where ink traces are faint and difficult to distinguish from the substrate. One key finding was that a trace-based labeling strategy, which focused strictly on visible ink traces without reconstructing missing portions, produced clearer results than the approach that attempted to extrapolate incomplete letters. While the latter method increased the amount of labeled

**Table 1**
Progression of Confident Letter Identifications Across Iterations and Segments

| Segment | Iteration 0 | Iteration 1 | Iteration 2 | Iteration 3 | New letters |
|---|---|---|---|---|---|
| 20241030152031 | 78 | 84 | 99 | 105 | 27 |
| 20241028121112 | 27 | 32 | 35 | 38 | 11 |
| 20241028121111 | 17 | 25 | 27 | 31 | 14 |
| 20241030083650 | 14 | 15 | 13 | 18 | 4 |
| 20241105112301 | 13 | 12 | 17 | 23 | 10 |
| 20241113090990 | 10 | 16 | 16 | 18 | 8 |
| 20241113070770 | 6 | 7 | 8 | 8 | 2 |
| 20241113080880 | 4 | 7 | 9 | 14 | 10 |
| 20241025062040 | 0 | 31 | 36 | 42 | 42 |
| 20241108111522 | 0 | 18 | 20 | 22 | 22 |
| 20241108120732 | 0 | 13 | 16 | 20 | 20 |
| 20241108120731 | 0 | 9 | 10 | 11 | 11 |
| 20241025150211 | 0 | 6 | 7 | 8 | 8 |
| 20241025145341 | 0 | 5 | 5 | 5 | 5 |
| 20241102160330 | 0 | 6 | 3 | 5 | 5 |
| **TOTAL** | **169** | **286** | **321** | **368** | **199** |

data, it introduced additional uncertainty and noise, reducing the clarity of detected text. Additionally, our results reinforce the value of expert verification in ensuring historical and linguistic accuracy, preventing misinterpretation of detected letterforms.

Challenges remain, particularly in distinguishing ink from the surrounding papyrus, but recent efforts have expanded the availability of labeled data. With our iterative approach, we have contributed to an improved dataset, which can now serve as a foundation for future model training. Building on this, future work should explore semi-supervised learning and active learning techniques to further optimize the annotation process while maintaining label accuracy. Additionally, enhancing preprocessing methods, such as contrast optimization and structure-preserving augmentation, may improve ink visibility and model robustness.

Beyond the technical aspects, collaboration between computer scientists, historians, and papyrologists will be essential for making meaningful progress in virtual unwrapping and text reconstruction. As deep learning continues to advance, these interdisciplinary efforts will play a critical role in recovering and interpreting texts from the Herculaneum scrolls and other fragile manuscripts, bringing lost historical knowledge back to light.

## Acknowledgments

This work uses data provided by the Vesuvius Challenge. We gratefully acknowledge Vesuvius Challenge team and the Scroll Prize organizers for providing access to the scanned Herculaneum scroll datasets. These datasets, including high-resolution XPCT scans and segmentation outputs, were made available through https://scrollprize.org and are central to the analyses presented in this paper. All the data used in the preparation of this paper were obtained from the EduceLab-Scrolls dataset [18].

We thank the entire Vesuvius Challenge community for their contributions to open research and their continued efforts in making ancient history accessible through modern technology.

## Online Resources

All relevant data used in this paper that exceeds the scope of the Vesuvius Challenge repository is openly accessible in our public GitHub repository https://github.com/FrankDeinzer/ki2025_data. Researchers are invited to reproduce or extend our pipeline.

## Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References


[1] The vesuvius challenge: Deciphering the secrets of the herculaneum papyri, 2024. URL: https://scrollprize.org/, accessed: 2025-06-30.

[2] G. Bertasius, H. Wang, L. Torresani, Is space-time attention all you need for video understanding?, arXiv preprint arXiv:2102.05095 (2021). URL: https://arxiv.org/abs/2102.05095.

[3] W. B. Seales, C. S. Parker, M. Segal, E. Tov, P. Shor, Y. Porath, From damage to discovery via virtual unwrapping: Reading the scroll from en-gedi., Science Advances 2 (2016) e1601247. URL: https://www.science.org/doi/abs/10.1126/sciadv.1601247. doi:10.1126/sciadv.1601247. arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.1601247.

[4] E. Brun, M. Cotte, J. Wright, M. Ruat, P. Tack, L. Vincze, C. Ferrero, D. Delattre, V. Mocella, Revealing metallic ink in herculaneum papyri, Proceedings of the National Academy of Sciences 113 (2016) 3751–3754. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1519958113. doi:10.1073/pnas.1519958113. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1519958113.

[5] C. Colini, O. Hahn, O. Bonnerot, S. Steger, Z. Cohen, T. Ghigo, T. Christiansen, M. Bicchieri, P. Biocca, M. Krutzsch, I. Rabin, The quest for the mixed inks, in: Second International Conference on Natural Sciences and Technology in Manuscript Analysis, Hamburg, 2016, pp. 41–48.

[6] P. Tack, M. Cotte, S. Bauters, E. Brun, D. Banerjee, W. Bras, C. Ferrero, D. Delattre, V. Mocella, L. Vincze, Tracking ink composition on herculaneum papyrus scrolls: Quantification and speciation of lead by x-ray based techniques and monte carlo simulations, Scientific Reports 6 (2016) 20763. doi:10.1038/srep20763.

[7] N. Labbé, D. Harper, T. Rials, T. Elder, Chemical structure of wood charcoal by infrared spectroscopy and multivariate analysis, Journal of agricultural and food chemistry 54 (2006) 3492–7. doi:10.1021/jf053062n.

[8] S. V. Vassilev, D. Baxter, L. K. Andersen, C. G. Vassileva, An overview of the composition and application of biomass ash. part 1. phase–mineral and chemical composition and classification, Fuel 105 (2013) 40–76. URL: https://www.sciencedirect.com/science/article/pii/S0016236112007545. doi:https://doi.org/10.1016/j.fuel.2012.09.041.

[9] S. Li, T. Zhang, J. Li, L. Shi, X. Zhu, J. Lü, Y. Li, Stabilization of pb(ii) accumulated in biomass through phosphate-pretreated pyrolysis at low temperatures, Journal of Hazardous Materials 324 (2017) 464–471. URL: https://www.sciencedirect.com/science/article/pii/S0304389416310184. doi:https://doi.org/10.1016/j.jhazmat.2016.11.014.

[10] G. Cavallo, M. Capasso, T. Dorandi, Libri scritture scribi a Ercolano: introduzione allo studio dei materiali greci, number Bd. 13 in Cronache ercolanesi, supplemento, Macchiaroli, 1983. URL: https://books.google.de/books?id=nAwjAQAAMAAJ.

[11] Y. Nader, The ink detection journey of the vesuvius challenge, 2024. URL: http://youssefnader.com/2024/02/06/the-ink-detection-journey-of-the-vesuvius-challenge/, accessed: 2025-06-30.

[12] ScrollPrize, volume-cartographer: A pipeline for segmentation and virtual unwrapping of herculaneum papyri, 2024. URL: https://github.com/ScrollPrize/volume-cartographer, accessed: 2025-06-30.

[13] J. Schillinger, V. Challange, Thaumato anakalyptor: Technical report and roadmap, 2024. URL: https://github.com/ScrollPrize/villa/blob/main/thaumato-anakalyptor/documentation/ThaumatoAnakalyptor___Technical_Report_and_Roadmap.pdf, accessed: 2025-06-30.

[14] S. Stabile, F. Palermo, I. Bukreeva, D. Mele, V. Formoso, R. Bartolino, A. Cedola, A computational

platform for the virtual unfolding of herculaneum papyri, Scientific Reports 11 (2021) 1695. URL: https://www.nature.com/articles/s41598-020-80458-z. doi:10.1038/s41598-020-80458-z.
[15] W. B. Seales, J. Griffioen, R. Baumann, M. Field, Analysis of herculaneum papyri with x-ray computed tomography, in: International Conference on nondestructive investigations and microanalysis for the diagnostics and conservation of cultural and environmental heritage, volume 7, e-Journal of Nondestructive Testing, Florence, Italy, 2011. URL: https://www.ndt.net/?id=11086, Special Issue 2011-10, ISSN 1435-4934.
[16] D. Baum, N. Lindow, H.-C. Hege, V. Lepper, T. Siopi, F. Kutz, K. Mahlow, H.-E. Mahnke, Revealing hidden text in rolled and folded papyri, Applied Physics A 123 (2017) 171. URL: https://link.springer.com/article/10.1007/s00339-017-0808-6.
[17] Y. Nader, Vesuvius challenge 2023 grand prize winner code repository, 2023. URL: https://github.com/younader/Vesuvius-Grandprize-Winner, accessed: 2025-06-30.
[18] S. Parsons, C. S. Parker, C. Chapman, M. Hayashida, Educelab-scrolls: Verifiable recovery of text from herculaneum papyri using x-ray ct, arXiv preprint arXiv:2304.02084 (2023). URL: https://arxiv.org/abs/2304.02084. doi:10.48550/arXiv.2304.02084, revised version v4 as of May 2024.
[19] Weights & Biases, 2025. URL: https://wandb.ai, accessed: 2025-06-30.
[20] S. E. M. Thompson, An Introduction to Greek and Latin Palaeography, Clarendon Press, Oxford, 1912. URL: https://archive.org/details/b29010408/page/144/mode/2up, enlarged edition of the author's Handbook of Greek and Latin Palaeography (3rd ed., 1906).
[21] V. Mocella, E. Brun, C. Ferrero, D. Delattre, Revealing letters in rolled herculaneum papyri by x-ray phase-contrast imaging, Nature communications 6 (2015) 5895. URL: https://doi.org/10.1038/ncomms6895. doi:10.1038/ncomms6895.