

## Master Plan

*July 25th, 2025*

### Vesuvius Challenge – Stage Two

- In 2023, Vesuvius Challenge [made a breakthrough](#), extracting more than four passages of never-before-seen text from inside an unopened (and unopenable) carbonized scroll. We have proven techniques for virtually unwrapping the papyrus scroll and recognizing the ink using machine learning. It wasn't clear it was possible until we did it.
- In 2024, we proved that the results from 2023 were not a unicum by [extracting words](#) from a second unopened carbonized scroll, while we focused on making the [virtual unwrapping](#) more efficient and automated.
- In 2025, (as of July 2025) while still automating unwrapping, the focus shifted to optimizing the scanning protocol, with the aim of X-ray imaging and reading the 300 extant scrolls, mostly in Naples. By pioneering cutting-edge X-ray scanning techniques—such as tetra-, hexa-, and octa-helical acquisitions—on the [BM18 beamline at ESRF](#), we can now scan a large scroll at 9.2  $\mu\text{m}$  pixel size in under two hours. This more than halves the scan time compared to our previous sessions, while also improving effective resolution. More than thirty scrolls have been scanned under varied setups, altering both resolution and contrast mechanisms. So far, ink remains elusive in all our new data. Nevertheless, we extracted the title from one of the “readable” unopened scrolls, confirming that the author was [Philodemus of Gadara](#).

Two key technical problems remain: unwrapping at scale and ink identification.

- Unwrapping at scale
  - The current bottleneck is tracing the papyrus surface inside the scan of the scroll (we call this “unwrapping”, and also “segmentation”). Currently we use semi-automated tracing aided by manual refinement. This method is still too time-consuming and affected by mistakes. Full manual tracing, as in 2023, can provide more accurate results at the expense of an even longer tracing process.
  - Full scrolls are 10cm-20cm wide and up to 15 meters long. With current techniques, it could cost \$1-5 million to unwrap an entire scroll. Given that there are 300 scrolls that need to be read, it could cost hundreds of millions or even more to unwrap all of them. Clearly impractical. Also, there are parts

of the scrolls that are so compressed, current techniques cannot unwrap them at all.

- A breakthrough is essential. We believe it's possible to bring the cost of unwrapping an entire scroll to \$5000 or below. It might even be possible to fully automate it.
- Our approach to solving unwrapping at scale will be to continue to leverage the community through a series of open source “progress prizes” we will award throughout the year, while hiring in full-time or part-time roles the most productive contributors to do the needed deep work.
- Ink identification
  - The ink-detection breakthroughs of 2023 and 2024 do not appear to generalize to the entire collection. Why is that? We have evidence that the ink signal enhanced by the machine learning model in the two readable scrolls corresponds either to large morphological cracks or to brighter spots under X-ray, indicating a higher concentration of metals. Unfortunately, ink-detection models trained to amplify these signals are not currently performing well on the other scrolls. As a result, we are now looking for different ink characteristics to train our models on.
  - Analysis of [fragments](#) suggests that when ink is neither metal-rich nor “cracked,” higher-resolution scans may be required to enhance the amount of signal in the data. We are currently (July 2025) exploring this route.

Improvements in both unwrapping at scale and ink identification are the main goals of 2025.

### **Vesuvius Challenge – Stage Three**

- Once unwrapping will be entirely automated, and ink identification will generalize across the full collection, we'll need to systematize and staff the scanning, segmenting, and reading pipeline.
- We expect that scanning and reading all the remaining 300 scrolls in the collection can be done in 2-3 years, depending on what we learn about the maximum speed of scanning with a protocol that also maximizes ink signal in the data.

### **Vesuvius Challenge – Stage Four**

- The final stage of the Vesuvius Challenge is inspiring the continued excavation of the Villa dei Papiri, and recovering in full the only surviving library from the ancient

world. It is a near-certainty that there are more scrolls waiting for us in the dirt. Perhaps just a few, but there could be thousands of them.

- Excavation is very expensive, but we expect this to be largely a political effort. Our hope is that the output of stages two and three above – previously unseen books from antiquity – will catalyze the will necessary to begin digging. If it does not, however, we will do whatever we can to make it happen.

### **Costs**

- We believe Stage 2 will cost \$5-6M. Thanks to many individual donors and to a generous donation of \$2,084,000 from the Musk Foundation this stage is partially funded.
- If an efficient particle accelerator protocol can be devised, we believe Stage 3 will cost \$4-8M. If it doesn't, Stage 3 will cost \$15M+, depending on the cost of beam time we are able to negotiate.

### **The Payoff**

- Overfit stories of history get rewritten
- Beautiful ancient literature is revealed
- A new renaissance of the classics
- Eternal glory