

# Ancient scrolls are being 'read' by machine learning—with human knowledge to detect language and make sense of them

March 13 2024, by C. Michael Sampson

---



‘An Eruption of Vesuvius,’ by Johan Christian Dahl (1824). (The Metropolitan Museum of Art), [CC BY](#)

A groundbreaking announcement for the recovery of lost ancient literature was recently made. Using a non-invasive method that harnesses [machine learning](#), an international trio of scholars retrieved 15 columns of ancient Greek text from within a carbonized papyrus from [Herculaneum](#), a seaside Roman town eight kilometers southeast of Naples, Italy.

Their achievement earned them a US\$700,000 grand prize from the [Vesuvius Challenge](#). The challenge sought to incentivize [technological development](#) by inviting [public participation](#) in the research.

It emerged from collaboration between computer scientist Brent Seales—who has [a long-standing interest](#) in non-invasive [technologies for studying](#) manuscripts—and technology investors Nat Friedman and Daniel Gross.

While the developments are exciting, technology is only part of the progress of scholarship. The work of reading and analyzing the new Greek and Latin texts recovered from the papyri will fall to human beings.

## **Buried in ash**

Like Pompeii, [Herculaneum](#) was buried by the catastrophic eruption of Mount Vesuvius in 79 CE.

Much of the ancient town remains underground. But in 1752, excavation uncovered hundreds of papyrus scrolls in the library of an elaborate Roman villa. The Herculaneum papyri [are the largest surviving example of an](#) intact ancient library preserved in the archaeological record: the library was found as it actually existed in 79 CE.

The precise number of books is unknown, says Michael McOsker, a research fellow in papyrology at University College London, and different methods of estimating give different results.

## **Carbonized papyri**

Starved of oxygen, the intense heat of Vesuvius' [pyroclastic flow](#) carbonized (but did not ignite) the papyri. Resembling lumps of coal to the eye, 18th-century excavators did not immediately recognize them as ancient books.

The papyri are so brittle that many were destroyed by early attempts to access their texts. Studying them has therefore always required ingenuity. In 1754, a [conservator and priest at the Vatican library](#) devised a machine for slowly unrolling them.

More recently, [multispectral photography](#) has dramatically improved their legibility. But until now, a [non-invasive method](#) that would leave the scrolls intact remained out of reach. Its development marks a significant breakthrough.

McOsker notes there are 659 items in the catalogue listed as "not unrolled," but some of these are parts of scrolls.

## **Sparking innovation**

To kick-start the challenge, Seales [made public](#) an array of high-resolution X-ray computed tomography (CT) scans of two scrolls as well as similar scans of detached fragments with visible ink. The latter are essential as a reference point (or "control") for innovative approaches.

The competition's design encouraged transparency and collaboration:

data published in the pursuit [of smaller goals](#) benefited all competitors. Additionally, transparency enabled the independent verification of results. Teams coalesced around shared ideas and approaches to the problem.

## **Text mentions music, taste, sight**

The challenge made news in [October](#), when the first letters were read: πορφυρας (a noun or adjective involving "purple").

By the end of 2023, the criteria for awarding the grand prize were met: four passages of 140 characters, with 85 percent of the letters recovered. [A Ph.D. student studying machine learning, an engineer studying computer science and a robotics student](#) were declared the victors.

According to McOsker, the text they retrieved mentions music twice, as well as the senses of taste and sight. He thinks it is likely a work about sensation and decision-making, in the tradition of [the philosopher Epicurus \(341–270 BCE\)](#). The challenge's papyrological team is still analyzing it.

## **Hundreds of rolls to be studied**

This year brings with it new goals: after five percent of one scroll was read in 2023, the challenge set a [2024 grand prize goal](#) of reading 90 percent of four scrolls. With hundreds of rolls yet to be studied, the new method of recovering the contents of the Herculaneum papyri is only getting started.

But several obstacles remain. The production of scans at sufficiently high resolution can't be done via ordinary equipment, but requires access to a facility with a particle accelerator. Access to the right equipment is

limited and costly. To date, four scrolls and numerous detached fragments [have been processed at a facility](#) near Oxford, England.

Most of the unopened scrolls are housed in Naples, and getting them safely to a facility will be complicated, as will reserving and paying for the beam time required to scan them.

Another limitation is that the technology for unrolling and flattening out a papyrus by virtual means—a process the challenge calls "segmentation"—is slow and expensive. Via current techniques, which involve a fair bit of manual manipulation, fully segmenting one scroll would cost US\$1–5 million. Segmentation needs to become much more efficient to avoid a bottleneck.

## **Critical minds needed**

Technology is only part of the equation. Essential to the challenge's work is an international team of papyrologists. Their role is to analyze the model's output of legible ancient Greek—and in so doing determine which approaches are most effective.

Papyrology is thrilling work, but also challenging and painstaking. It requires mastery of ancient languages and ideas as well as the puzzle-solver's ability to fill in the inevitable gaps. Papyrology is a niche specialization: in the larger world of classics, papyrologists are rare birds. The number of Herculaneum specialists is even fewer.

For the challenge truly to succeed, we're going to need critical minds as well as whizbang technology. There's potentially a fair bit of new ancient philosophy headed our way, but it needs to be pieced together into a coherent text—letter by letter, word by word, sentence by sentence—before it can be studied more widely. That's going to require scholars.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Ancient scrolls are being 'read' by machine learning—with human knowledge to detect language and make sense of them (2024, March 13) retrieved 29 April 2024 from <https://phys.org/news/2024-03-ancient-scrolls-machine-human-knowledge.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.