All CCS and Cross-sectoral Cultural Heritage & Museums Research / Surveys Artificial Intelligence Heritage

# Cracking the code: can AI help us decipher ancient languages?

CREATIVESUNITE

Co-funded by the European Union

While artificial intelligence accelerates the processing of lost ancient scripts like Linear A by rapidly matching patterns across thousands of fragments, expert linguists emphasize that true decipherment still requires human historical insight and comparative anchors.

Jane Adkins, Dublin City University

Every ancient language that has ever been deciphered needed an anchor. Usually that's a bilingual text, like the Rosetta Stone, or a known relative to compare it to. Linear A, the writing system of the Bronze Age Minoan civilisation on the Greek island of Crete, has neither.

Linear A is described as a "language isolate" because it has no confirmed link to any known language, living or dead. As such, it has posed a significant challenge for linguists over the past century.

Etruscan, a language of Italy that was used before the rise of the Roman Empire, has fared only slightly better. A partial vocabulary has been collected from short funerary inscriptions, but its grammar and deeper meaning still elude linguists.

Both languages are exactly the kind of puzzle that AI has recently been pulled into – not by replacing the linguists, but by supercharging their hunches. So does AI actually stand a chance of finishing what a century of human effort hasn't?

A recent case shows both the appeal and the catch. In June 2026, a self-taught AI engineer and amateur linguist claimed a breakthrough on Linear A. It started from a single guess that one unknown word in a prayer inscription derived from a Semitic root meaning "to dwell" or "to inhabit".

He then used AI-built programming scripts to test that sound-pattern against the rest of a collected Linear A corpus. He was reportedly able to assign values to 40 signs and compiled a 408-word lexicon, showing – in his view – that Linear A is an extinct member of the Semitic language group, which encompasses languages such as Hebrew and Aramaic.

Linear A inscription on a clay tablet from Hagia Triada, Crete, probably dating to the 15th century BC. From the Archaeological Museum of Heraklion. Zde, CC BY-SA

## Scant evidence

It's worth being precise about what AI did and didn't do here. It didn't have the idea. The engineer did.

The AI was the research assistant. It rapidly cross-checked a human hypothesis against thousands of characters, work that would have taken a human months by hand. That division of labour is the pattern currently worth paying attention to, far more than any single claim that is still under review.

So where does AI actually help, and where does it hit a wall? It's genuinely strong at large-scale pattern testing. It can check whether a hunch about one sign or word holds up across an entire archive in minutes instead of years. It can spot repeated sequences a human eye would miss, and restore damaged or fragmentary inscriptions by predicting likely missing characters.

AI also does something distinct called "cross-lingual transfer", where a model trained on a known language can sometimes infer patterns in a closely related unknown one. This is similar to how knowing Spanish helps you guess at Portuguese.

Researchers have already used this kind of approach successfully on scripts such as Ugaritic, where the language family is known. Ugaritic is an extinct Semitic language, spoken during the late Bronze Age (about 1300-1190BC) in the ancient coastal city of Ugarit (in what is now Syria).

But there's a hard ceiling: statistical pattern matching can't manufacture meaning out of nothing. It needs an anchor – a known language family, or a bilingual text – to tell it which patterns are meaningful and which are coincidence.

Linear A and Etruscan are hard precisely because that anchor is missing or incomplete, and no amount of computing power

invents one from scratch.

With a large enough corpus, an AI model could hypothetically get fluent enough to converse with a Minoan or Etruscan speaker on the model's own statistical terms. This means that it could recognise repeating patterns of words and other structural units of a language that would allow the AI model to repurpose the terms in a different context. One that might approximate a limited conversation, at least on the surface.

But the AI system still wouldn't be able to hand a human a translation, because fluency and meaning aren't the same thing here. The model can learn which signs follow which, and which words cluster together, without ever knowing what any of them actually refer to.

This is also why verifying any AI-assisted claim about these languages is so difficult, and it's a problem bigger than any one case. Normally, you'd check a proposed translation or interpretation against native speakers, other texts or expert consensus built up over decades. None of that exists for a genuinely undeciphered language. Short of a time machine, there's no way to check.

Small surviving corpora make this worse: Linear A's entire surviving corpus is about 7,500 characters, short enough to fit on a single screen and with that little data, almost any hypothesis can find scattered matches to support it.

That is why claims in this space lean so heavily on independent expert scrutiny and peer review rather than statistical confidence scores. It is also why "AI found a pattern" and "AI found the correct meaning" are very different claims that are easy to blur together.

None of this means that AI is a dead end for these languages. Quite the opposite: it's a real accelerant, able to compress years of manual cross-referencing into mere minutes and let more people attempt these problems than institutional resources ever allowed.

But it doesn't remove the two ingredients that decipherment has always required. One is a genuine comparative anchor. The other is rigorous human review, to tell a real breakthrough from an appealing coincidence.

Until one of those shows up for Linear A or Etruscan, AI's role in deciphering languages will stay what it is today: a very fast assistant to a very old, very human puzzle.

Jane Adkins, PhD candidate, School of Computing, Dublin City University
Image: An inscription in Etruscan hangs above the heads of a noble couple in this wall painting from the Tomb of the Shields, Tarquinia, Italy. Prisma Archivo