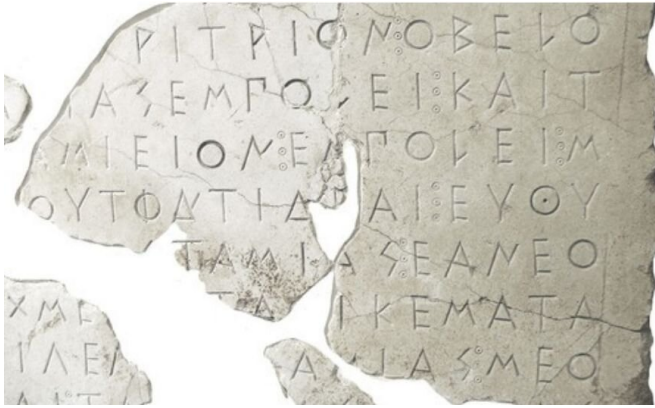


# Deep learning enlightens scholars puzzling over ancient texts

20 October 2019, by Nancy Cohen



Damaged inscription: a decree concerning the Acropolis of Athens (485/4 BCE). IG I 3 4B. (CC BY-SA 3.0, Wikimedia)

Deep learning can help scholars restore ancient Greek texts. Specifically, researchers at University of Oxford (Thea Sommerschild and Professor Jonathan Prag) and DeepMind (Yannis Assael) built Pythia, training a neural network to guess missing words or characters from Greek inscriptions.

These were on surfaces including stone, ceramic and metal. They were between 1500 and 2600 years old. *New Scientist* [reported](#) that AI beat humans in deciphering damaged tablets.

"In a head-to-head test, where the AI attempted to fill the gaps in 2949 damaged inscriptions, human experts made 30 percent more mistakes than the AI. Whereas the experts took 2 hours to get through 50 inscriptions, Pythia gave its guesses for the entire cohort in seconds."

Starting out, the authors knew that restoring text was a time-consuming task—even for expert epigraphists. They set out to evaluate the difficulty of the restoration task at hand—and thereby judge

the impact of our work—with the help of two doctoral students with epigraphical expertise. The scholars were allowed to use the training set to search for "parallels."

Gege Li wrote on Friday in *New Scientist*. The AI seems to be better than humans at filling in missing words, but this is no Team A versus Team B competition. Rather, the AI technique, [said](#) Li, "may be most useful as a collaborative tool, where researchers use it to narrow down the options."

Many ancient inscriptions have become eroded or damaged over the centuries. The authors said that "Only a small minority of surviving inscriptions are fully legible and complete."

With segments of text lost, how could one try to fill in the blanks of missing words? As Li said, it would mean looking at the rest of the inscription and looking at other similar texts.

Consider *New Scientist's* report on what the AI, dubbed Pythia, was able to do: (1) Pythia learned to recognise patterns in 35,000 relics, with over 3 million words. (2) Patterns it picks up on include the context in which different words appear, the grammar, and also the shape and layout of inscriptions.

The accomplishment is reflected in the title of their paper which now up on arXiv: "Restoring ancient text using [deep learning](#): a [case study](#) on Greek epigraphy."

To aid the epigraphist, Pythia doesn't just give the scholar a single prediction. Rather, it returns multiple predictions as well as the level of confidence for each result.

"Specifically, we provide a set of the Top 20 predictions decoded using beam search." With 20 suggestions to fill the gap, it is up to the person to

select the best one. "It's all about how we can help the experts," said Assael. To be sure, their position is that Pythia can serve as an assistive method in digital epigraphy.

deep learning: a case study on Greek epigraphy? arXiv:1910.06262 [cs.CL] [arxiv.org/abs/1910.06262](https://arxiv.org/abs/1910.06262)

© 2019 Science X Network

[Encyclopaedia Britannica](#): Epigraphy is "the study of written matter recorded on hard or durable material. The authors similarly provided a definition. They stated that "Epigraphy is the study of documents, 'inscriptions', written on a durable surface (stone, ceramic, metal) by individuals, groups and institutions of the past."

The team talked about Pythia's future potential, and they pointed out that it is the combination of machine learning and epigraphy that has the potential to impact meaningfully the study of inscribed textual cultures.

"By open-sourcing PYTHIA, and PHI-ML's processing pipeline, we hope to aid future research and inspire further interdisciplinary work."

Why their research matters: Pythia, they wrote, is "the first ancient text restoration model that recovers missing characters from a damaged [text](#) input using [deep neural networks](#)." The authors believe that Pythia "sets the state-of-the-art in [ancient text](#) restoration."

[Faculty of Classics](#) at the University of Oxford site similarly commented on Pythia's strengths. "The architecture works at both the character- and word-level, thereby effectively handling long-term context information, and dealing efficiently with incomplete word representations. This makes it applicable to all disciplines dealing with ancient texts (philology, papyrology, codicology) and applies to any language (ancient or modern)."

The Faculty of Classics at the University of Oxford said that an online Python notebook, Pythia, and PHI-ML's processing pipeline have been open sourced on [GitHub](#).

With origins in London in 2010, [DeepMind](#), meanwhile, is in the frontlines of artificial intelligence research.

**More information:** Restoring ancient text using

APA citation: Deep learning enlightens scholars puzzling over ancient texts (2019, October 20) retrieved 25 October 2020 from

<https://techxplore.com/news/2019-10-deep-enlightens-scholars-puzzling-ancient.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.*