

1. Shah, A. K. & LaForest, M. *Nature* **603**, 297–301 (2022).
2. Levay, K. E., Freese, J. & Druckman, J. N. *Sage Open* **6**, <https://doi.org/10.1177/2158244016636433> (2016).
3. Mullinix, K. J., Leeper, T. J., Druckman, J. N. & Freese, J. *J. Exp. Polit. Sci.* **2**, 109–138 (2015).
4. National Academies of Sciences, Engineering, and Medicine. *Proactive Policing: Effects on Crime and*

- Communities* (National Academies Press, 2018).
5. Liberman, N., Trope, Y. & Stephan, E. *Social Psychology: Handbook of Basic Principles* 2nd edn (eds Kruglanski, A. W. & Higgins, E. T.) 353–381 (Guilford, 2007).
  6. Mussweiler, T. *Psychol. Rev.* **110**, 472–489 (2003).
  7. Hershfield, H. E., John, E. M. & Reiff, J. S. *Policy Insights Behav. Brain Sci.* **5**, 209–215 (2018).
  8. van Gelder, J. L., Luciano, E. C., Weulen Kranenbarg, M. & Hershfield, H. E. *Criminology* **53**, 158–179 (2015).
  9. Silver, E. & Miller, L. L. *Criminology* **42**, 551–584 (2004).

The authors declare no competing interests.

This article was published online on 2 March 2022.

## Machine learning

# Mind the gap as AI guesses at lost Greek inscriptions

Charlotte Roueché

The use of artificial intelligence (AI) is transforming many areas of research. A new AI tool helps to fill in missing text and estimate the timeframe and geographical origin of ancient inscriptions. See p.280

The possibility that artificial intelligence (AI) will automate tasks and thus destroy certain jobs is advancing steadily into more and more areas of life; the waves are now lapping even on the quiet shores of ancient-world studies. On page 280, Assael *et al.*<sup>1</sup> present an AI tool called Ithaca, which has been developed to bring deep learning to the world of classical studies and the interpretation of transcribed ancient Greek texts, which were originally inscribed on stone. But this advance should not be interpreted as a threat to centuries of tradition – rather as a complement to them.

The study of the past is always based on insufficient evidence, and the more distant the past, the more fragmentary the evidence. Historians regularly use hypothesis to bridge

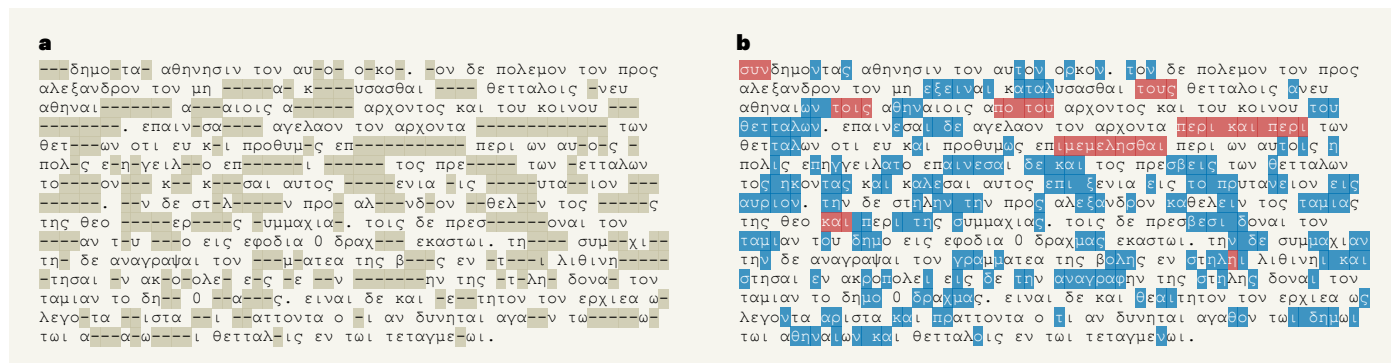
the gaps, and regularly modify their hypotheses. In the case of the study of inscribed texts (epigraphy), this is literally true. Epigraphers have to develop skills, based on painstaking study and wide experience, to fill the gaps in texts that survive only in fragments, and to assess the material's date and provenance. Until now, this work has usually relied on the deep experience of a community of scholars, who gradually learn to identify the linguistic conventions of particular societies and their customary procedures (such as ratifying a decree). Ithaca offers a computer-based way to approach these tasks (Fig. 1), and uses deep learning to replicate and improve the results. Such a method can provoke real fears that human understanding might become

superfluous – fears not so different from the concerns felt in other fields of expertise.

In the area of humanities, scholars of the classical world have been surprisingly advanced in their adoption of digital tools with which to present and exploit their materials. Several early resources, such as the rich collection of literary texts in the Perseus Digital Library of ancient Greek and Latin texts (see [go.nature.com/3t5yvub](http://go.nature.com/3t5yvub)), were conceived in the 1980s, made available in compact-disc format in the 1990s, and then transferred to the Internet during the 2000s. The early birth of such projects meant that many were designed and understood principally as tools for searching. Since then, the Internet has been used to present data that would otherwise be prohibitively expensive to publish – such as the ground-breaking online publication of the Vindolanda Tablets (see [go.nature.com/3jz5c1d](http://go.nature.com/3jz5c1d)), a set of correspondence between Roman soldiers and their families who were based around Hadrian's Wall in Britain in the late first and early second centuries AD.

This early adoption of digital formats inspired and facilitated the development of Ithaca by providing an easily accessible training set for this AI tool. The principal resource used by Assael and colleagues is the Searchable Greek Inscriptions data set of the Packard Humanities Institute in Santa Clarita, California. This is not a random collection of material, but a presentation of 178,551 transcribed texts to which scholarly judgement has already been applied, and both legible words and letters and illegible spaces have been carefully reported. Another enabling tool was the Lexicon of Greek Personal Names at the University of Oxford, UK, a project conceived well before the Internet was born, based on decades of careful scholarship assigning locations and dates to each example of a name used in the ancient Greek world. Ithaca has harnessed this expertise and extends it, mimicking the neural processes of the scholar.

Use of these resources allowed Ithaca to



**Figure 1 | Predicting missing text in an ancient Greek inscription.** **a**, Scholars previously transcribed this fragmented inscription, noting gaps where letters are missing (highlighted dashes). **b**, Assael *et al.*<sup>1</sup> report the development of an artificial-intelligence tool called Ithaca, which is trained to suggest text that might fill these sorts of gaps. The

text proposed by Ithaca matches the suggestions previously made by scholars (text highlighted in blue), with a few exceptions (text highlighted in red). The high level of accuracy achieved by Ithaca indicates how it can enhance and enrich the work of scholars. (Figure based on Extended Data Fig. 4 of ref. 1.)

## News & views

learn the prevailing patterns among personal names and in the language of inscribed texts from particular places and periods. The authors then applied this tool to assess a group of fragmentary texts, and to produce suggestions for the restoration of missing text, as well as for provenance and date. When these outputs were compared against the corresponding judgements of expert scholars, the authors found that Ithaca provided a notably higher level of accuracy and a richer result.

The Vindolanda correspondence was written on perishable strips of wood, and survived by chance. In areas of the Graeco-Roman world where good stone was easily available, public and private documents were often inscribed on this more-substantial material. Some survive in reasonably complete form, but the passage of time has rendered many more into fragments. The discovery and publication of such texts, beginning in the early nineteenth century, has transformed our engagement with the ancient past. Careful work of this kind, for example, continues to modify our understanding of the world of classical Athens – a community that produced many documents, and that was located among hills offering an abundance of excellent marble on which such documents could easily be inscribed.

As a result, the study of the ancient world, which was initially focused almost entirely on

literary texts, has expanded to engage with a wide range of sources, from shopping lists to poetry. Inscriptions can throw up new information at any moment, illuminating domestic life, the economy or politics, and requiring constant reassessment of what we think we know. Such challenging data might come from newly excavated discoveries (which is what grabs headlines), or from reanalysis of material we thought we understood, using tools such as Ithaca.

**“Ithaca acts like a bloodhound, finding the trail for the scholar, but not making the final decisions.”**

Ithaca’s potential for discovery is clear; it is not about displacing diagnostic expertise, but turbo-charging it. It is easy for our hopes and expectations to skew our vision, but this tool approaches any fragment of text without human prejudice. It can work consistently across texts, whether they are of evident importance or seemingly insignificant. Then, crucially, it offers not a fixed answer for what the text corresponding to missing gaps might be, but rather a range of answers ranked by their probability. The stated aim of the project

is “to maximize the collaborative potential between historians and deep learning”. The scholar is not displaced; instead, Ithaca acts like a bloodhound, finding the trail for the scholar, but not making the final decisions.

We use the language of teaching and learning when describing the development of an AI tool. Teaching forces us to analyse and explain what we do; this is precisely the process we engage in when we aim to pass on our skills and accumulated knowledge to the next generation. Ithaca has been built on the training set that scholarship provides, and extends the existing principles of analysis. Confronting and working with such a diagnostic tool will produce important new knowledge, but it should also help scholars to have a better understanding of their own intellectual processes. The use of AI should not render the scholar redundant, but should instead challenge their understanding of what they thought they knew.

**Charlotte Roueche** is in the Department of Classics, King’s College London, London WC2R 2LS, UK.  
e-mail: [charlotte.roueche@kcl.ac.uk](mailto:charlotte.roueche@kcl.ac.uk)

1. Assael, Y. et al. *Nature* **603**, 280–283 (2022).

The author declares no competing interests.

nature  
briefing

# Back to the lab

## How to work smarter in the new normal

Sign up for a free short email series with top tips, resources and advice for those starting or returning to research, brought to you by the Nature Briefing team.

[go.nature.com/backtothelab](https://go.nature.com/backtothelab)

A108920

