Article

Restoring and attributing ancient texts using deep neural networks

https://doi.org/10.1038/s41586-022-04448-z

Received: 16 August 2021

Accepted: 19 January 2022

Published online: 9 March 2022

Open access

Check for updates

Yannis Assael[1,6]✉, Thea Sommerschield[2,3,6]✉, Brendan Shillingford[1], Mahyar Bordbar[1], John Pavlopoulos[4], Marita Chatzipanagiotou[4], Ion Androutsopoulos[4], Jonathan Prag[5] & Nando de Freitas[1]

Ancient history relies on disciplines such as epigraphy—the study of inscribed texts known as inscriptions—for evidence of the thought, language, society and history of past civilizations[1]. However, over the centuries, many inscriptions have been damaged to the point of illegibility, transported far from their original location and their date of writing is steeped in uncertainty. Here we present Ithaca, a deep neural network for the textual restoration, geographical attribution and chronological attribution of ancient Greek inscriptions. Ithaca is designed to assist and expand the historian's workflow. The architecture of Ithaca focuses on collaboration, decision support and interpretability. While Ithaca alone achieves 62% accuracy when restoring damaged texts, the use of Ithaca by historians improved their accuracy from 25% to 72%, confirming the synergistic effect of this research tool. Ithaca can attribute inscriptions to their original location with an accuracy of 71% and can date them to less than 30 years of their ground-truth ranges, redating key texts of Classical Athens and contributing to topical debates in ancient history. This research shows how models such as Ithaca can unlock the cooperative potential between artificial intelligence and historians, transformationally impacting the way that we study and write about one of the most important periods in human history.

Epigraphy is the study of texts—inscriptions—written directly on durable materials (stone, pottery, metal) by individuals, groups and institutions of the ancient world[2,3]. Thousands of inscriptions have survived to our time, but many have been damaged over the centuries and their texts are now fragmentary. Inscriptions may also be moved or trafficked far from their original location[4], and radiocarbon dating is unusable owing to the inorganic nature of most inscribed supports. Specialist epigraphers must then reconstruct the missing text, a process known as text restoration (Fig. 1), and establish the original place and date of writing, tasks known as geographical attribution and chronological attribution, respectively[5]. These three tasks are crucial steps towards placing an inscription both in history and within the world of the people who wrote and read it[6,7]. However, these tasks are non-trivial, and traditional methods in epigraphy involve highly complex, time-consuming and specialized workflows.

When restoring damaged inscriptions, epigraphers rely on accessing vast repositories of information to find textual and contextual parallels[8]. These repositories primarily consist of a researcher's mnemonic repertoire of parallels and, more recently, of digital corpora for performing 'string matching' searches. However, differences in the search query can exclude or obfuscate relevant results, and it is almost impossible to estimate the true probability distribution of possible restorations. Attributing an inscription is equally problematic—if it was moved, or if useful internal dating elements are missing, historians must find alternative criteria to attribute the place and date of writing (such as letterforms, dialects)[9]. Inevitably, a high level of generalization is often involved (chronological attribution intervals can be very long).

Deep learning for epigraphy

Here we overcome the constraints of current epigraphic methods by using state-of-the-art machine learning research. Inspired by biological neural networks, deep neural networks can discover and harness intricate statistical patterns in vast quantities of data[10]. Recent increases in computational power have enabled these models to tackle challenges of growing sophistication in many fields[11–14], including the study of ancient languages[15–18].

We present Ithaca, a deep neural network architecture trained to simultaneously perform the tasks of textual restoration, geographical attribution and chronological attribution. Ithaca, which was named after the Greek island that eluded the hero Odysseus' homecoming, was trained on inscriptions written in the ancient Greek language and across the ancient Mediterranean world between the seventh century BC and the fifth century AD. This choice was due to two main reasons. First, the variability of contents and context of the Greek epigraphic record, which makes it an excellent challenge for language processing; and second, the availability of digitized corpora for ancient Greek, an essential resource for training machine learning models.

[1]DeepMind, London, UK. [2]Department of Humanities, Ca' Foscari University of Venice, Venice, Italy. [3]Center for Hellenic Studies, Harvard University, Washington, DC, USA. [4]Department of Informatics, Athens University of Economics and Business, Athens, Greece. [5]Faculty of Classics, University of Oxford, Oxford, UK. [6]These authors contributed equally: Yannis Assael, Thea Sommerschield. ✉e-mail: yannisassael@google.com; thea.sommerschield@unive.it

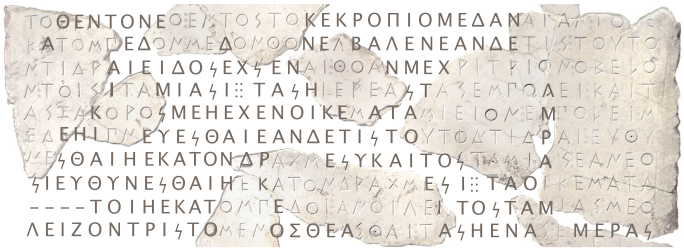


Fig. 1 | Restoration of a damaged inscription. This inscription (*Inscriptiones Graecae*, volume 1, edition 3, document 4, face B (*IG* I[3] 4B)) records a decree concerning the Acropolis of Athens and dates to 485/4 BC. Marsyas, Epigraphic Museum, WikiMedia CC BY 2.5.

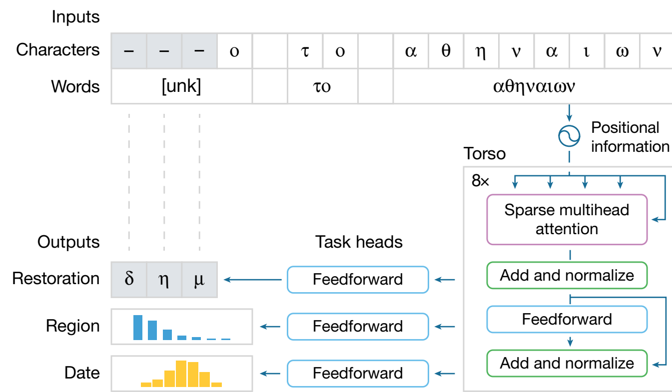


Fig. 2 | Ithaca's architecture processing the phrase 'δήμο το αθηναίων' ('the people of Athens'). The first three characters of the phrase were hidden and their restoration is proposed. In tandem, Ithaca also predicts the inscription's region and date.

Working with Greek inscriptions

To train Ithaca, we developed a pipeline to retrieve the unprocessed Packard Humanities Institute (PHI)[19,20] dataset, which consists of the transcribed texts of 178,551 inscriptions. This process required rendering the text machine-actionable, normalizing epigraphic notations, reducing noise and efficiently handling all irregularities. Each PHI inscription is assigned a unique numerical ID, and is labelled with metadata relating to the place and time of writing. PHI lists a total of 84 ancient regions; whereas the chronological information is noted in a wide variety of formats, varying from historical eras to precise year intervals, written in several languages, lacking in standardized notation and often using fuzzy wording[21]. After crafting an extended ruleset to process and filter the data (Methods), the resulting dataset I.PHI is to our knowledge the largest multitask dataset of machine-actionable epigraphical text, containing 78,608 inscriptions.

Ithaca is a model for epigraphic tasks

The architecture of Ithaca was carefully tailored to each of the three epigraphic tasks, meaningfully handling long-term context information and producing interpretable outputs to enhance the potential for human–machine cooperation. To begin, contextual information is captured more comprehensively by representing the inputs as words; however, parts of words could have been lost over the centuries. To address this challenge, we process the input text as character and word representations jointly, representing damaged, missing or unknown words with a special symbol '[unk]'.

Next, to enable large-scale processing, Ithaca's torso is based on a neural network architecture called the transformer[22], which uses an attention mechanism to weigh the influence of different parts of the input (such as characters, words) on the model's decision-making process. The attention mechanism is informed of the position of each part of the input text by concatenating the input character and word representations with their sequential positional information. Ithaca's torso consists of stacked transformer blocks: each block outputs a sequence of processed representations of which the length is equal to the number of input characters, and the output of each block becomes the input of the next. The final output of the torso is passed to three different task heads that handle restoration, geographical attribution and chronological attribution, respectively. Each head consists of a shallow feedforward neural network, specifically trained for each task. In the example shown in Fig. 2, the restoration head predicts the three missing characters; the geographical attribution head classifies the inscription among 84 regions; and the chronological attribution head dates it to between 800 BC and AD 800.

Interpreting the outputs

Our intention was to maximize the collaborative potential between historians and deep learning. Ithaca's architecture was therefore designed to provide intelligible outputs, while featuring multiple visualization methods to augment the interpretability of the model's predictive hypotheses. For the task of restoration, instead of providing historians with a single restoration hypothesis, Ithaca offers a set of the top 20 decoded predictions ranked by probability (Fig. 3a). This first visualization facilitates the pairing of Ithaca's suggestions with historians' contextual knowledge, therefore assisting human decision-making. This is complemented by saliency maps, a method used to identify which unique input features contributed the most to the model's predictions, for both the restoration and attribution tasks (Fig. 3d and Extended Data Fig. 5a).

For the geographical attribution task, Ithaca classifies the input text among 84 regions, and the ranked list of possible region predictions is visually implemented with both a map and a bar chart (Fig. 3b). Finally, to expand interpretability for the chronological attribution task, instead of outputting a single date value, we predict a categorical distribution over dates (Fig. 3c). By so doing, Ithaca can handle ground-truth labels more effectively, as the labels correspond to date intervals. More precisely, Ithaca discretizes all dates between 800 BC and AD 800 into 10-year bins, resulting in 160 decades. For example, the date range 300–250 BC is represented as 5 decades of equal 20% probability, whereas an inscription dated to 305 BC would be assigned to the single-decade-bin 300–310 BC with 100% probability.

Experimental evaluation

To compare performance in the three epigraphic tasks, we use four methods. First, we evaluate the difficulty of the restoration task by assigning two evaluators with epigraphical expertise ('ancient historian') a set of damaged inscriptions to restore, using the training set to search for textual parallels. Second, we provide the human experts with a ranked list of Ithaca's top 20 restoration hypotheses to inform their predictions ('ancient historian and Ithaca'), therefore assessing the true impact of our work as a cooperative research aid. Third, as a computational baseline we reimplement our previous work Pythia[15]—a sequence-to-sequence recurrent neural network for the task of ancient-text restoration. Finally, for the attribution tasks, we introduce an ablation of the epigrapher's workflow, the 'onomastics' baseline: annotators were tasked with attributing a set of texts, exclusively using the known distribution of Greek personal names across time and space to infer geographical and chronological indicia[27].

We introduce the following metrics to measure each method's performance. For restoration, to obviate the lack of ground truths in damaged inscriptions, we artificially hide 1 to 10 characters of undamaged input text and treat the original sequences as the target. The first metric used is the character error rate (CER), which counts the normalized

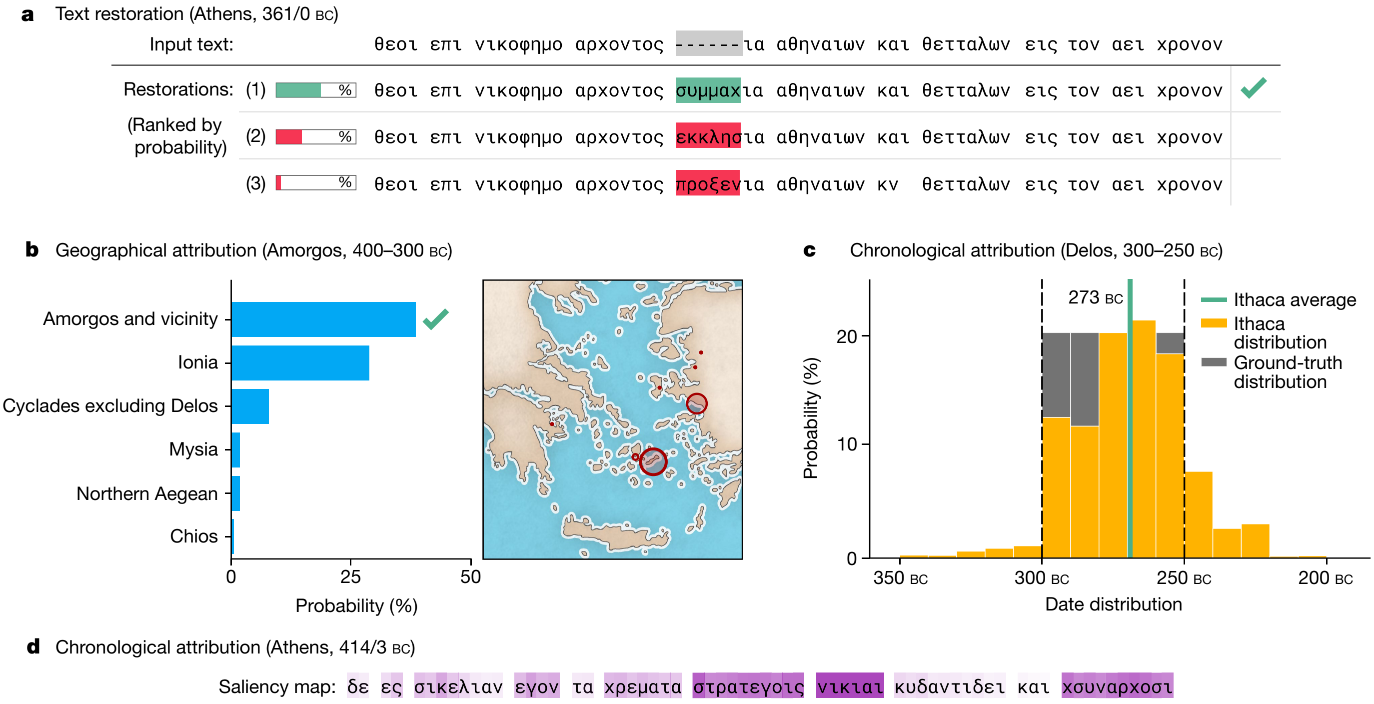


Fig. 3 | Ithaca's outputs. a, Restoration predictions for six missing characters (dashes) in an Athenian inscription (*IG* II^2 116). The top restoration, in green, is correct (συμμαχία, 'alliance'). Note how the following hypotheses (ἐκκλησία, 'assembly'; and προξενία, 'treaty between state and foreigner'), highlighted in red, typically occur in Athenian political decrees[23], revealing Ithaca's receptivity to context. **b**, Geographical attribution of an inscription from Amorgos (*IG* XII 7, 2). Ithaca's top prediction is correct, and the closest predictions are neighbouring regions. **c**, Date distribution for an inscription from Delos (*IG* XI 4, 579). The ground-truth date interval 300–250 BC is shown in grey; Ithaca's predicted distribution is shown in yellow and has a mean at 273 BC (green). Ithaca's predictions show a higher confidence for the interval's higher date margin, therefore potentially narrowing the broad ground-truth dating bracket. **d**, Chronological attribution saliency map for an Athenian inscription (*IG* I^3 371). The colour intensity illustrates the importance of each input. Ithaca focuses on the personal name (Νικίας, 'Nikias') and the Greek commanders' rank (στρατεγοίς, 'generals'). Nikias had a key role in the great Athenian expedition to Sicily[24–26], the historical event to which this very inscription pertains. Ithaca dates the inscription to 413 BC, matching the exact range proposed by historians (414–413 BC).

differences between the top predicted restoration sequence and the target sequence. Furthermore, we use top-*k* accuracy to measure whether the correct restoration or region label for geographical attribution is among the top *k* predictions, therefore quantifying Ithaca's potential as an assistive tool. For chronological attribution, we use a distance metric (Methods) to measure the distance in years from the predictive distribution's mean and the ground-truth interval, the latter being defined by a minimum and a maximum date.

As shown in Table 1, for the task of restoration, Ithaca consistently outperforms the competing methods, scoring a 26.3% CER and 61.8% top 1 accuracy. Specifically, our model achieves a 2.2× lower (that is, better) CER compared with human experts, whereas Ithaca's top 20 predictions achieve a 1.5× improved performance compared with Pythia, with an accuracy of 78.3%. Notably, when pairing historians with Ithaca (ancient historian and Ithaca), human experts achieve an 18.3% CER and 71.7% top 1 accuracy, therefore demonstrating a considerable 3.2× and 2.8× improvement compared with their original CER and top 1 scores. Regarding the attribution to regions, Ithaca has 70.8% top 1 and 82.1% top 3 predictive accuracy. Finally, for chronological attribution, whereas the onomastics human baseline predictions are within an average of 144.4 and median of 94.5 years from the ground-truth date intervals, Ithaca's predictions, based on the totality of texts, have an average distance of 29.3 years from the target dating brackets, with a median distance of only 3 years.

Table 1 | Experimental results

	Restoration			Region		Date
Method	CER (%)	Top 1 (%)	Top 20 (%)	Top 1 (%)	Top 3 (%)	Years
Ancient historian and Ithaca	**18.3**	**71.7**				
Ithaca	26.3	61.8	78.3	**70.8**	**82.1**	**29.3**
Pythia	47.0	32.6	53.9			
Ancient historian	59.6	25.3				
Onomastics				21.2	26.5	144.4

Evaluating methods for text restoration, geographical attribution (region) and chronological attribution (date) on I.PHI's test set of n=7,811 inscriptions. For 'CER' and 'years', lower scores are better. For 'top 1', 'top 3' and 'top 20', higher scores are better. For each metric, the best performing method is in bold.

Contributing to historical debates

Our experimental evaluation effectively demonstrates Ithaca's impact on the study of inscriptions, and their consequent value as historical evidence. First, Ithaca can discover epigraphic patterns on an unprecedented scale and in unparalleled detail, harnessing substantial quantities of epigraphic data (I.PHI) to achieve the high performance observed in all three epigraphic tasks. Moreover, whereas Ithaca may have outperformed historians in the first baseline, the combination of a historian's own (contextual) knowledge alongside Ithaca's assistive input resulted in an even greater improvement over the model's performance. This collaborative potential is augmented by Ithaca's design decisions, and by the different visualization aids increasing the interpretability of outputs, therefore enabling historians to evaluate multiple hypotheses.

As a consequence, Ithaca could help historians narrow the wide or vague date brackets they are sometimes forced to resort to, by helping increase precision and establish relative datings for historical events, and even contributing to current methodological debates in ancient history.

Indeed, to demonstrate Ithaca's creative potential, we applied our model to a contemporary dispute concerning the dating of a group of inscriptions whose interpretation is central to the political history of classical Athens. Historians disagree on whether these decrees should pre- or post-date 446/5 BC depending on the (dis)belief in using specific letterforms as dating criteria (the three-bar sigma dating convention)[28]. In recent years, the validity of this dating convention was called into question[29]—the dates of many decrees have been pushed to the 420s BC, therefore profoundly influencing our understanding of Athenian imperialism[30].

This group of disputed Athenian decrees exists in our dataset: their dating labels follow the conventional 'higher' dates (pre-446/5 BC). We excluded these texts from the dataset and trained Ithaca on all of the remaining inscriptions. Notably, Ithaca's predictions for these held-out texts independently align with the most recent dating breakthroughs, therefore overturning the conventional historical reading based on the sigma dating criterion. More specifically, whereas the I.PHI labels are on average 27 years off the 'lower' dating proposed by modern re-evaluations, Ithaca's predictions are on average only 5 years off the newly proposed ground truths.

This example eloquently illustrates how models such as Ithaca can contribute to key methodological debates on the chronological reorganization of Athenian imperialism, one of the most important moments in Greek history. In no instance do Ithaca's predictions for this group of inscriptions exceed 433 BC: Ithaca's average predicted date for all of these decrees is 421 BC. Historians may now use Ithaca's interpretability-augmenting aids (such as saliency maps) to examine these predictions further and bring more clarity to Athenian history.

Conclusions

Ithaca is to our knowledge the first epigraphic restoration and attribution model of its kind. By substantially improving the accuracy and speed of the epigrapher's pipeline, it may assist the restoration and attribution of newly discovered or uncertain inscriptions, transforming their value as historical sources and helping historians to achieve a more holistic understanding of the distribution and nature of epigraphic habits across the ancient world. To achieve this goal, our interdisciplinary team created an open-source and publicly available interface (https://ithaca.deepmind.com), enabling historians to use Ithaca for their personal research, while facilitating its development for further applications.

In fact, the methods introduced in this research apply to all disciplines dealing with ancient text (papyrology, numismatics, codicology), to any language (ancient or modern), also integrating additional metadata (inscription images, stylometrics). Furthermore, Ithaca's quintessentially interactive nature as a cooperative research aid lends itself as an effective set-up for future machine learning research by adding humans into the training loop.

In conclusion, the transformational impact of this work lies in delivering state-of-the-art research aids that extend the scope of ancient history and the humanities.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04448-z.

1. Davies, J. & Wilkes, J. Epigraphy and the Historical Sciences (British Academy, 2012).
2. Osborne, R. In *The Oxford History of Historical Writing: Volume 1: Beginnings to AD 600* (eds Feldherr, A. & Hardy, G.) 97–121 (Oxford Univ. Press, 2011).
3. Bodel, J. P. Epigraphic Evidence: Ancient History from Inscriptions (Routledge, 2001).
4. Tsirogiannis, C. The itinerary of a stolen stele. *UNESCO Cour.* **4**, 18–20 (2020).
5. Bruun, C. & Edmondson, J. C. in *The Oxford Handbook of Roman Epigraphy* (eds Bruun, C. & Edmondson, J. C.) 13–20 (Oxford University Press, 2015).
6. Macmullen, R. The epigraphic habit in the Roman empire. *Am. J. Philol.* **103**, 233–246 (1982).
7. Nawotka, K. Epigraphic Culture in the Eastern Mediterranean in Antiquity (Routledge, 2021).
8. Osborne, R. & Rhodes, P. J. Greek Historical Inscriptions 478-404 BC xvii–xviii (Oxford Univ. Press, 2017).
9. Cooley, A. The Cambridge Handbook to Latin Epigraphy 398–434 (Cambridge Univ. Press, 2012).
10. Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning (MIT Press, 2016).
11. Brown, T. B. et al. Language models are few-shot learners. In *Proc. Advances in Neural Information Processes (NeurIPS)* Vol. 33 (eds Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H.) 1877–1901 (Curran Associates, 2020).
12. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
13. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
14. Silver, D. et al. Mastering the game of go without human knowledge. *Nature* **550**, 354–359 (2017).
15. Assael, Y., Sommerschield, T. & Prag, J. Restoring ancient text using deep learning: a case study on Greek epigraphy. In *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* 6368–6375 (Association for Computational Linguistics, 2019).
16. Bamman, D. & Burns, P. J. Latin BERT: a contextual language model for classical philology. Preprint at https://arXiv.org/abs/2009.10053 (2020).
17. Kang, K. et al. Restoring and mining the records of the Joseon dynasty via neural language modeling and machine translation. In *Proc. 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)* 4031–4042 (Association for Computational Linguistics, 2021).
18. Shen, T., Quach, V., Barzilay, R. & Jaakkola, T. Blank language models. In *Proc. 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (eds Webber, B., Cohn, T., He, Y. & Liu, Y.) 5186–5198 (Association for Computational Linguistics, 2020).
19. Packard Humanities Institute. The Packard Humanities Institute's Searchable Greek Inscriptions (2005); https://inscriptions.packhum.org/
20. Gawlinski, L. Review: Packard Humanities Institute's Searchable Greek Inscriptions (2017); https://classicalstudies.org/scs-blog/laura-gawlinski/review-packard-humanities-institutes-searchable-greek-inscriptions
21. Iversen, P. A. The Packard Humanities Institute (PHI) Greek epigraphy project and the revolution in Greek epigraphy. *Abgadiyat* **2**, 51–55 (2007).
22. Vaswani, A. et al. Attention is all you need. In *Proc. Advances in Neural Information Processing Systems (NeurIPS)* Vol. 30 (eds Guyon, E. et al.) 5998–6008 (Curran Associates, 2017).
23. Hedrick, C. W. Jr Democracy and the Athenian epigraphical habit. *Hesperia* **68**, 387–438 (1999).
24. Wesley, E. T. A new restoration of *I.G.* I^2 297. *Class. Q.* **14**, 230–231 (1964).
25. Thucydides. 6.31.
26. Kagan, D. The Peace of Nicias and the Sicilian Expedition (Cornell University Press, 1991).
27. Parker, R. Data in Online Database "Lexicon of Greek Personal Names (LGPN)" (Univ. Oxford, 2019).
28. Rhodes, P. After the three-bar sigma controversy: the history of Athenian imperialism reassessed. *Class. Q.* **58**, 500–506 (2008).
29. Mattingly, H. B. The Athenian Empire Restored: Epigraphic and Historical Studies 1–4 (Univ. Michigan Press, 1996).
30. Ma, J., Papazarkadas, N. & Parker, R. Interpreting the Athenian Empire (Duckworth, 2009).

Methods

Previous work

In recent years, several works have proposed traditional machine learning approaches to the study of ancient texts. This body of work has focused on optical character recognition and visual analysis[31–34], writer identification[35–37] and text analysis[38–44], stylometrics[45] and document dating[46]. It is only very recently that scholarship has begun to use deep learning and neural networks for optical character recognition[47–55], text analysis[56], machine translation of ancient texts[57–59], authorship attribution[60,61] and deciphering ancient languages[62,63], and been applied to study the form and style of epigraphic monuments[64].

The closest work to Ithaca is our 2019 research on ancient text restoration: Pythia[15]. Pythia was to our knowledge the first ancient text restoration model to use deep neural networks, and was followed by blank language models[18], Babylonian[65] and Korean text translation and restoration[17], Latin BERT for language modelling, part-of-speech tagging, word sense disambiguation and word similarity[16], and the classification of Cuneiform tablets by period[66].

Ithaca is to our knowledge the first model to tackle the three central tasks in the epigrapher's workflow holistically. Not only does it advance the previous state-of-the-art set by Pythia, but it also uses deep learning for geographical and chronological attribution for the very first time and on an unprecedented scale. Ithaca offers interpretable outputs, showcasing the rising importance of cooperation between human experts and machine learning[67]—as exemplified by our experimental evaluation.

Most importantly, this work shows how matching human experts with deep learning architectures to tackle tasks collaboratively can surpass the individual (unaided) performance of both humans and model on the same tasks. Indeed, recent medical research[68,69] further confirms the importance of hybrid architectures in addressing real-world problems. The present work makes human expert interaction possible by visualizing the output probability distributions for all tasks using multiple charts and maps, and augmenting their interpretability by means of saliency maps. It is our hope that this work may set a new standard for the field of digital epigraphy, by using advanced deep learning architectures to support the work of ancient historians.

Generating the I.PHI corpus

When restoring damaged inscriptions, epigraphers conjecture the total number of missing characters based on grammatical and syntactical considerations, and on the reconstructed physical form of the text[5]. Conjectured missing characters that cannot be restored are conventionally marked with periods or hyphens, one hyphen equating to one missing character. Moreover, PHI presents interpretive transcriptions of the texts (including capitalization, punctuation, word division, lower-case letter conversion).

Thus, moving from the PHI dataset, we substantially expand the ruleset for filtering human annotations previously conceived for Pythia, rendering the text machine-actionable. We removed 9,441 duplicate texts and filtered out all inscriptions under 50 characters in length, whereas, in Pythia's dataset, we had excluded all texts with fewer than 100 characters. To increase the amount of available text, we retained the supplements proposed by epigraphers (conventionally added between square brackets), and we matched the number of unrestored characters with an equal number of '–' symbols, as is commonly done by epigraphers (Extended Data Fig. 1).

Each PHI inscription is assigned to a region of the ancient Mediterranean world (Extended Data Fig. 2), and includes an additional metadata string referring to the date proposed by epigraphers for the text (Extended Data Fig. 1). The chronological information is noted in a variety of formats (historical eras, precise year intervals); in several languages (including Latin); ranging before (BCE) and after (CE) the Common Era; lacking in standardized notation ('early', 'first half', '1st half', 'beginning', 'beg.') and often using fuzzy wording ('late 7th/6th ac.', 'ca. 100 a.?', 'bef. 64 AD'). After crafting an extended ruleset, we succeeded in generating well-defined date intervals for 60% of all PHI inscriptions, as the chronological metadata of the remaining 40% is either missing or unprocessable. The resulting I.PHI dataset contains 1.93× more inscriptions than the previous Pythia's dataset. The texts of which the numerical PHI identifier (PHI ID) ended in 3 or 4 were held out and used as test and validation sets, respectively (Extended Data Table 1).

Ithaca architecture

Inputs. For each inscription, the input of the model consists of (1) a sequence of character embeddings (real-valued vectors, each representing the character of the alphabet that occurs at the corresponding position of the inscription); (2) an equally long sequence of word embeddings (real-valued vectors, each representing the vocabulary word at the corresponding character position of the inscription; Fig. 2); and (3) positional embeddings (also real-valued vectors, each representing a position of the input sequence). The first two kinds of embeddings are randomly initialized and learned when training Ithaca (via backpropagation). The positional embeddings are also trainable and they are initialized with a separate sinusoidal function per dimension[22] to maintain a symmetrical distance between neighbouring steps and smoothly decay over the maximum length of 768 characters. Our vocabulary includes every word appearing more than 10 times in I.PHI (35,884 words), while damaged or 'unknown' (under-represented) words are rendered with an '[unk]' symbol. The joint use of character and word embeddings enables the architecture of Ithaca to be both character- and context-aware[70–72]. Finally, the input sequence is padded with a start-of-sentence character '<'.

Torso. The three input sequences are combined by concatenating the different embeddings per-character position and the resulting sequence is fed through the torso of the model. The architecture of Ithaca's torso consists of eight stacked transformer decoder blocks, inspired by the large-scale transformer model BigBird[73]. Every block uses four sparse attention heads (using global, local and random attention mechanisms), which reduce the context-length dependency from quadratic to linear, therefore enabling the model to handle lengthier sequences[73] compared with classical transformers. Furthermore, the attention mechanism is 'multi-head' (Fig. 2) in the sense that it can learn to consider different types of information extracted from the input. For example, different attention heads may be sensitive to particular character sequences, or more perceptive to certain words and phrases with distinctive morphosyntactic or semantic features. Finally, to overcome problems that hinder the stacking of such complicated blocks, each transformer block uses residual connections and layer normalization (shown as 'add and normalize' in Fig. 2).

Task heads. Ithaca's torso outputs a sequence whose length is equal to the number of input characters, and each item in this sequence is a 2,048-dimensional embedding vector. Each task head consists of a two-layer feedforward network followed by a softmax function. There are three different task heads, handling region attribution, chronological attribution and restoration respectively. To predict the regions and dates, Ithaca uses the first output embedding ($t = 1$) and passes it on to the two corresponding heads. This arrangement is similar to that of DocBERT[74] and works better than other pooling methods (such as mean- and max-pooling over the output embeddings) in our experimental evaluation. Finally, for the restoration task, Ithaca uses the remaining output embeddings ($t > 1$) as there is a direct correspondence with the input text characters: for each missing character position, the corresponding output embedding of the torso is fed to the head of the restoration task, which predicts the missing character.

Data preparation and augmentation

I.PHI may be the first multitask dataset of machine-actionable epigraphical text, but its size is still several orders of magnitude smaller than modern typical language datasets. To avert the risk of overfitting, which is common in large-scale deep neural network architectures, we apply several data augmentation methods, described below, to artificially increase the size of I.PHI's training set. Our preliminary experimental evaluation found that these methods are crucial in achieving the reported performance. These augmentation methods are applied anew whenever a training inscription is re-encountered in each training epoch.

Text clipping. For each inscription, we select an arbitrary section of its text and ignore the remaining text. We implement this by first sampling a segment length between 50 and 768 characters, and then sampling the starting index of the segment. This method helps Ithaca to generalize and improve the handling of partial inputs.

Text masking. Forcing the model to rely on contextual information often leads to improvements in prediction. To achieve this in our model, during training, we randomly hide up to half of the input text by replacing sequences of characters sampled from a geometric distribution ($P = 0.1$) with '–'. This span masking is intended to replicate the distribution over the length of missing characters estimated from the dataset, and uses the hidden ground-truth characters as target labels for the restoration task.

Word deletion. During training, we also delete words from each input text (without replacing them with any special characters in this case) with a 20% probability. Here, the goal is again to increase variability in the training data to improve the model's ability to generalize over all possible ways in which inscriptions are damaged[75].

Sentence swap. By randomly swapping sentences in the input text with a 25% probability, we generate multiple input–label pairs for the auxiliary task of next-sentence prediction (NSP)[75] (see below).

Data circularity

Ithaca's source dataset (PHI) is a synthesis of generations of scholarly research. Epigraphers typically restore texts and attribute them chronologically through a process of induction. Textual restorations are proposed on the basis of parallels, mediated by wider historical and linguistic knowledge; chronological attributions are proposed partly from archaeological and contextual information, partly from textual form and content, and partly from textual and material parallels. The texts on which Ithaca trains include previous scholarly restorations; and the dates recorded are the product of accumulated scholarly knowledge and induction from archaeological, historical and textual study. This might be thought to imply circularity, but that would be true only if Ithaca were operating in a world of objective data and aiming to offer a single objectively true solution. Rather, Ithaca is an assistive tool aiming to improve on and facilitate a scholarly process of induction, model uncertainty and propose possible solutions for the scholar to consider.

Considering textual restoration, Ithaca avoids the risk of 'history from square brackets'[76–78] (assuming any proposed restoration to be ground truth, meaning the accepted consensus, rather than merely one of several hypotheses), because none of Ithaca's proposed restorations are assumed to be objectively certain—instead, they are presented as plausible suggestions. Furthermore, the inclusion of existing scholarly conjectures within the training set itself does not constitute a form of 'history from square brackets', as such conjectures are themselves plausible restorations achieved by a process of induction and considered acceptable by one or more experts, and as such are precisely the sort of result that Ithaca itself aims to generate. The value of Ithaca is indeed its ability to learn from the largest possible dataset of attested and possible texts, making the underlying process of inductive reasoning as powerful as possible, and so generating possible restorations for scholars to evaluate.

As for chronological attribution, the dataset on which Ithaca trains is founded in the past study of multiple elements (such as archaeological provenance, material form, textual content and form). Ithaca in turn learns through close attention to the text alone. The attributions proposed by Ithaca therefore have their basis in the inductive study of a vast textual dataset and its correlation to chronological data that are more broadly derived. Ithaca is therefore able to bring some refinement to those attempts to date the texts through the application of machine learning specifically to the textual patterns in that data. Thus, Ithaca is, in this case, a part of that scholarly process, and no more or less circular in its reasoning than any other scholar.

Training on epigraphic tasks

For the task of restoration, we use the text-masking augmentation method to mask parts of the input and produce ground truths. We subsequently use a cross-entropy loss to train Ithaca to predict the missing characters. The cross-entropy loss is also used for geographical attribution, using the region metadata as target labels. We further apply label smoothing with a coefficient of 10% to avoid overfitting and to provide historians with a smoother distribution of predicted hypotheses. For the task of chronological attribution, Ithaca discretizes all dates between 800 BC and AD 800 with a bin size of 10 years. This range covers the majority of the PHI dataset entries and encompasses the conventional date range for Greek epigraphy. The processed ground-truth date intervals are discretized into bins of equal probability, forming the target probability distribution. The limitations of discretizing and amalgamating date ranges of different levels of precision based on past scholarship have been noted[79,80]—the scale of data on which Ithaca trains, together with the increased attention to textual patterns (compared with the previous paragraph), at least partially meet that challenge. We then use the Kullback–Leibler divergence to minimize the difference between target and predicted probability distribution (Fig. 3c).

Finally, to allow for better modelling of context, we introduce a next sentence prediction loss, an auxiliary function common to language modelling tasks[81]. During training, we randomly shuffle some of the sentences of the input text, and at the end of each (non-final) sentence (marked by a full stop, '.') we predict whether the next sentence is in the correct order (valid) or a product of the shuffling augmentation. By deploying the torso's output embeddings for the full stops, we introduce an additional feedforward network that uses binary cross-entropy to predict the validity of the next sentence whenever a '.' character appears.

Using this setup, Ithaca was trained for a week on 128 Tensor Processing Units (TPU) v4 pods on the Google Cloud Platform. The effective batch size was 8,192 texts and a LAMB optimizer[82] was used to optimize Ithaca's parameters with a learning rate of 3×10^{-4}. Using Bayesian optimization hyperparameter search, the loss functions of each task were combined using the following function:

$$L = 3 \times L_{\mathrm{Restoration}} + 2 \times L_{\mathrm{Region}} + 1.25 \times L_{\mathrm{Date}} + 0.01 \times L_{\mathrm{NSP}}.$$

We do not use a separate masked (token) language modelling loss, which is commonly used when pretraining language models, as it is very similar to the restoration loss, although the latter masks characters instead of tokens.

To obtain Ithaca's textual restoration predictions, we select a sequence of missing characters to predict and use Beam Search with a beam width of 100. Instead of using a standard sequential Beam Search, we take advantage of Ithaca's non-autoregressive nature[83–85], and use a non-sequential one instead. Each beam starts with the prediction scoring the highest confidence[86], then proceeds iteratively to restore at each time-step the characters of which the certainty is the highest. We found that this version of Beam Search performed substantially better in our evaluation metrics. For region attribution, the outputs are presented as a plot of the top 10 predictions; for chronological attributions, we visualize the model's predictive distribution over possible date bins. Finally, to reduce the variance of random segment selections, we repeat the process ten times and report results averaged over the iterations.

Ancient historian baseline

The evaluators for ancient text restoration were two graduate students of ancient history, with 7 years of historical and linguistic training and specializing in Greek history and epigraphic documents. Thus, they can be assumed to be more capable than the 'average' ancient historian, but not yet equivalent to (the very small number) of established specialists in the field. The scholars were allowed to use the training set to search for textual 'parallels', and made an average of 50 restorations in 2 h.

Although Ithaca can indeed propose restoration hypotheses faster, and model its prediction uncertainty, it cannot make choices on the basis of historical and material context. Thus, the experimental setup cannot be considered to be direct comparison between human historians and machine learning, nor are the evaluators assumed to be a proxy for all historians. Instead, the experiment was intended to measure the difficulty of the task and the potential for cooperative artificial intelligence.

Onomastics baseline

Greek nomenclature is commonly used by epigraphers as one of several elements to inform their attribution predictions[87]. Inspired by this method in the wider epigraphic workflow, we designed an 'onomastic' baseline, of which the predictions are based exclusively on the metadata associated with Greek personal names. Five annotators searched for name(s) appearing in a set of inscriptions in the Lexicon of Greek Personal Names (LGPN), a database recording the geographical and chronological distribution of ancient names[27], and based their attribution hypotheses on the LGPN's distribution data. Evaluators were also provided with the inscription's date or place of writing for the geographical or chronological attribution tasks, respectively.

Restoration metrics

To evaluate different restoration methods, for every inscription, we predict a sequence of 1–10 contiguous missing characters. These lengths account for 83% of the distribution of missing character lengths in I.PHI, and enable comparisons with both previous work and the human baselines. Note that, thanks to the text-masking augmentation adopted during training, Ithaca could potentially restore up to half of the input text.

Although the number of characters to be predicted reflects the difficulty of the task, the restored sequences in the test sets held out for human evaluation might not necessarily maintain the same distribution of lengths (as they were a subset of the test set). Thus, instead of reporting only the average scores over the entire test set (as done in previous work), we chose to account for these length discrepancies and compute the average scores for each restored sequence length. First, we computed a separate CER for all samples of each length (between 1–10 characters),

$$\mathrm{CER}_l = \frac{1}{\sum_i^N I_{\mathrm{len}_i=l}} \sum_i^N I_{\mathrm{len}_i=l} \times \frac{\mathrm{EditDistance}\left(\mathrm{pred}_i, \mathrm{target}_i\right)}{l},$$

where I is the indicator function, len_i denotes the length of the i-th sample, N is the number of samples, pred_i is the predicted sequence of missing characters of the i-th sample and target_i the corresponding target sequence. We next calculate the average for all lengths:

$$\mathrm{CER}_{\mathrm{score}} = \frac{1}{L} \sum_l^L \mathrm{CER}_l.$$

where $L = 10$ is the maximum length.

As human annotators annotated only a subset of the test set owing to time constraints, macro-averaging assigns equal importance to all sample lengths to represent the difficulty of the task independently of dataset statistics, and therefore enabling a fair comparison of the methods. Similarly, for accuracy, we first computed a separate accuracy per length, and then the average:

$$\mathrm{accuracy}_l = \frac{1}{\sum_i^N I_{\mathrm{len}_i=l}} \sum_i^N I_{\mathrm{len}_i=l} \times I_{\mathrm{pred}_i=\mathrm{target}_i},$$

$$\mathrm{accuracy}_{\mathrm{score}} = \frac{1}{L} \sum_l^L \mathrm{accuracy}_l.$$

Chronological attribution metric

As our model outputs a predictive distribution in the chronological attribution task, we introduce an interpretable metric to measure the distance in years between a prediction and the ground-truth interval (Fig. 3c). More specifically, we use a distance metric between the mean of the predictive distribution and the target ground-truth interval; the latter is defined by a minimum ($\mathrm{gt}_{\mathrm{min}}$) and a maximum ($\mathrm{gt}_{\mathrm{max}}$) date in years:

$$\mathrm{Years} = \begin{cases} 0, & \text{if } \mathrm{gt}_{\mathrm{max}} \geq \mathrm{pred}_{\mathrm{avg}} \geq \mathrm{gt}_{\mathrm{min}} \\ |\mathrm{pred}_{\mathrm{avg}} - \mathrm{gt}_{\mathrm{max}}|, & \text{if } \mathrm{pred}_{\mathrm{avg}} > \mathrm{gt}_{\mathrm{max}} \\ |\mathrm{pred}_{\mathrm{avg}} - \mathrm{gt}_{\mathrm{min}}|, & \text{if } \mathrm{pred}_{\mathrm{avg}} < \mathrm{gt}_{\mathrm{min}} \end{cases}.$$

Model selection

The final model was obtained by storing the best-performing model on the validation set by using a combined metric that sums the accuracy for textual restoration and geographical attribution, and the distance in years divided by 100 for chronological attribution to make the magnitude comparable. The extensive computational resources required to train our model made the Pareto frontier computation infeasible.

Chronological attribution results

Ithaca's predictions are 5× closer to ground truths than those recorded in the onomastics baseline (144.4 years). More specifically, Ithaca's average date prediction is within 28.7 years of the ground-truth date interval, and the median is only 3 years. The results are shown in detail in Extended Data Fig. 3.

Restoring full texts with Ithaca

To overcome memory constraints and length limitations for long inscriptions (>768 characters), Ithaca can be applied iteratively to restore all missing text in a damaged inscription. We experimented with this option on inscription *IG* II^2 116, which is missing 378 characters, and compared Ithaca's predictions with those of our previous work Pythia on the same text, using the authoritative edition published by Rhodes and Osborne as ground truths[88]. The models' correct restorations are highlighted in green (Extended Data Fig. 4), and the erroneous ones in red. In a real-world scenario, both Ithaca and Pythia would provide a ranked set of 20 restoration hypotheses. The comparison in performance between Pythia and Ithaca is stark

(74 versus 45 mistakes): moreover, in all cases in which the restoration is in red, the ground-truth sequence existed within the beam of Ithaca's top 20 hypotheses.

Geographical attribution of Delphic inscriptions

Epigraphers determine the original location where an inscription was written by examining the personal names, local or regional dialectal varieties, and idiosyncratic lexicon or style of an inscription. Moving from this methodological premise, and to discover underlying patterns in Ithaca's geographical predictions, we compute statistics to track the words that appear most frequently in texts whose region Ithaca predicts correctly. Thus, for each word of the test set, we compute an average accuracy and a frequency of appearance. This visualization is intended to evaluate whether the occurrence of particular words could be correlated to the model's geographical attributions.

The most frequent words that appear in texts with high prediction accuracy clustered primarily in inscriptions from the region of Delphi, and pertained to the epigraphic genre of 'manumission inscriptions' (Extended Data Table 2 for an example). Ancient Greek society depended heavily on unfree labour, but slaves could be freed through a process known as 'manumission', which was publicly documented and certified by inscriptions[89,90]. Over 1,000 such texts dating between around 201 BC and AD 100 have been found in Delphi[91,92]. The words appearing in Ithaca's accuracy statistics are identified as typical of these manumission texts, which are in turn distinctive of this region (for example, ἐπίστευσε, ἀποδμενος, καταδουλισμωι, βεβαιωτήρ, ωνάν): these words could therefore be underpinning the correct attribution predictions (a detailed example is offered in Extended Data Table 2). Further study can now be dedicated to investigating stylized manumissions as distinctive of Delphi.

To further assess the impact of Ithaca's output visualization techniques in a real-world scenario, we also analysed the saliency maps for geographical attribution of the manumission inscriptions. Indeed, the saliency maps for the Delphic inscription *BCH* 66/67 (1942/3) 82,9, for example, highlight words typically found in manumission texts and which also appear in Ithaca's word statistics: these words (ἐπίστευσε, ἐλευθερος, ποιέουσα, ἀποτρέχουσα) have the most important role in the geographical attribution of the inscription, while also betraying the text's genre as a typical slave manumission inscription (Extended Data Fig. 5b).

Redating disputed Athenian decrees

In the absence of helpful internal evidence of a text's date (for example, the mention of known historical figures[93]), epigraphers typically derive an approximate date on the basis of a text's content, letterforms and grammatical criteria. For example, one of the most notorious methodological debates in epigraphy concerns the 'three-bar sigma' dating convention, which holds that no Athenian public document containing the three-bar sigma letter (ϟ) could be dated after the year 446/5 BC, when the letter was supplanted by the four-bar sigma (Σ). On the basis of this chronological benchmark, a group of inscriptions whose interpretation is central to the political history of Classical Athens, and which feature the earlier letter ϟ, were dated to pre-446/5 BC by many authoritative corpora[28,94]. This set of decrees exists in the PHI dataset (Extended Data Table 3), and their dating labels follow the conventional 'higher' dating of the three-bar sigma criterion.

However, this orthodox dating system soon proved to be problematic: the high dates proposed for these decrees did not agree with contemporary literary accounts reporting on Athenian imperialist policies. Few historians contested the validity of the sigma criterion[29,95], but in 1990 photo-enhancement and laser scanning confirmed the down-dating of an inscription featuring the three-bar sigma (the Egesta decree, *IG* I³ 11) from 458 to 418 BC[96]. Over the following decade, the sigma's traditional cut-off date was revisited, and the dates of other decrees were also pushed back[28,97].

Ithaca's predictions for this set of disputed inscriptions independently align with the most recent dating breakthroughs (Extended Data Fig. 6). For example, the (in)famous Chalcis decree (*IG* I³ 40; Extended Data Fig. 7), which records an oath of allegiance sworn by the city of Chalcis to Athens[98] and traditionally dated to 446/5 BC[28], is attributed by Ithaca to 420 BC, therefore concurring with the lower dating hypothesis of 424/3 BC proposed by more recent scholarship[99]. Perhaps the most compelling example of Ithaca's prediction independently aligning with a lower dating hypothesis is the decree of Kleinias (*IG* I³ 34)[100], regulating the collection of tribute across the Athenian empire. The sigma dating system would assign the inscription to 448/7 BC[28], but scholars have recently challenged this orthodoxy and proposed the earlier date of 425/4 BC[101]. Ithaca's prediction agrees precisely with the latter, dating the famous decree to 424 BC.

Ithaca has re-dated a number of these key inscriptions with striking accuracy (Extended Data Table 3). Although it may seem slight, this 40/30-year chronological reorganization has considerable implications for our grasp of Athenian imperial behaviour, leading historians to a more profound understanding of one of the most momentous periods of ancient history[28,97]. The fact that Ithaca was trained on the largest available dataset of Greek epigraphic texts makes it possible to challenge or overcome individual biases or, indeed, errors in the existing academic tradition, notwithstanding the fact that the dataset in question is originally based on the accumulated academic tradition.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Ithaca was trained on The Packard Humanities Institute's Searchable Greek Inscriptions public dataset, PHI, which is available online (https://inscriptions.packhum.org/). The complete processing workflow for transforming the dataset to a machine-actionable format suitable for training Ithaca (I.PHI) is available at GitHub (https://github.com/sommerschield/iphi) under Apache License 2.0. The LGPN (https://www.lgpn.ox.ac.uk/) was used by annotators for the onomastics baseline to track the geographical and chronological distribution of ancient names. The PeriodO gazetteer (https://client.perio.do/) was used as a reference for mapping the PHI historical time periods to the chronological range metadata of I.PHI. The Pleiades gazetteer (https://pleiades.stoa.org/) was used as a reference for mapping the PHI region names to the geographical coordinates used in the geographical attribution map visualizations.

Code availability

Ithaca's training and inference source code is available at GitHub (https://github.com/deepmind/ithaca) under Apache License 2.0, along with the trained weights, licensed under Creative Commons Attribution-ShareAlike 4.0 International. A public interface for historians using Ithaca for their research (that is, restoration and attribution of Greek inscriptions, use of all visualization tools discussed in the present paper) is available online (https://ithaca.deepmind.com). Neural networks were developed with JAX v.0.2.9 (https://github.com/google/jax/), Flax v.0.3.0 (https://github.com/google/flax), and Haiku v.0.0.4 (https://github.com/deepmind/dm-haiku). The XLA compiler is bundled with JAX and does not have a separate version number. Dataset processing and analysis used Python v.3.7 (https://www.python.org/), NumPy v.1.19.2 (https://github.com/numpy/numpy), SciPy v.1.5.2 (https://www.scipy.org/), pandas v.1.1.3 (https://github.com/pandas-dev/pandas), beautifulsoup4 v.4.9.0 (https://www.crummy.com/software/BeautifulSoup/) and Google Colab (https://research.google.com/colaboratory), which is an online service and does not have a version number. Visualizations were generated using matplotlib v.3.4.2 (https://matplotlib.org/), seaborn v.0.11.1 (https://seaborn.pydata.org/) and GeoPandas v.0.9.0 (https://geopandas.org/).

31. Garz, A., Eichenberger, N., Liwicki, M. & Ingold, R. HisDoc 2.0: toward computer-assisted paleography. *Manuscr. Cult.* **7**, 19–28 (2015).
32. Shaus, A. Computer Vision and Machine Learning Methods for Analyzing First Temple Period Inscriptions. PhD thesis, Tel Aviv Univ. (2017).
33. Soumya, A. & Kumar, G. H. Classification of ancient epigraphs into different periods using random forests. In *Proc. 2014 Fifth International Conference on Signal and Image Processing* 171–178 (IEEE Computer Society, 2014).
34. Terras, M. & Robertson, P. Image to Interpretation: An Intelligent System to Aid Historians in Reading the Vindolanda Texts (Oxford Univ. Press, 2006).
35. Faigenbaum-Golovin, S. et al. Algorithmic handwriting analysis of Judah's military correspondence sheds light on composition of biblical texts. *Proc Natl Acad. Sci. USA* **113**, 4664–4669 (2016).
36. Panagopoulos, M., Papaodysseus, C., Rousopoulos, P., Dafi, D. & Tracy, S. Automatic writer identification of ancient Greek inscriptions. *Trans. Pattern Anal. Mach. Intel.* **31**, 1404–1414 (2009).
37. Tracy, S. V. & Papaodysseus, C. The study of hands on Greek inscriptions: the need for a digital approach. *A. J. Archaeol.* **113**, 99–102 (2009).
38. Koppel, M., Michaely, M. & Tal, A. Reconstructing ancient literary texts from noisy manuscripts. In *Proc. Fifth Workshop on Computational Linguistics for Literature (NAACL-HLT)* (eds Feldman, A., Kazantseva, A. & Szpakowicz, S.) 40–46 (Association for Computational Linguistics, 2016).
39. Lee, J. & Haug, D. Porting an ancient Greek and Latin treebank. In *Proc. Seventh International Conference on Language Resources and Evaluation (LREC)* (eds Calzolari, N. et al.) (European Language Resources Association, 2010).
40. Rao, R. P. et al. A Markov model of the Indus script. *Proc. Natl Acad. Sci. USA* **106**, 13685–13690 (2009).
41. Rao, R. P. et al. Entropic evidence for linguistic structure in the Indus script. *Science* **324**, 1165–1165 (2009).
42. Rao, R. P. et al. Entropy, the Indus script, and language: a reply to R. Sproat. *Comput. Linguist.* **36**, 795–805 (2010).
43. Vatri, A. & McGillivray, B. The Diorisis ancient Greek corpus: linguistics and literature. *Res. Data J. Hum. Soc. Sci.* **3**, 55–65 (2018).
44. Yadav, N. et al. Statistical analysis of the Indus script using *n*-grams. *PLoS ONE* **5**, e9506 (2010).
45. Gianitsos, E., Bolt, T., Chaudhuri, P. & Dexter, J. Stylometric classification of ancient Greek literary texts by genre. In *Proc. 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* 52–60 (Association for Computational Linguistics, 2019).
46. Baledent, A., Hiebel, N. & Lejeune, G. Dating ancient texts: an approach for noisy French documents. In *Proc. LT4HALA 2020 1st Workshop on Language Technologies for Historical and Ancient Languages* 17–21 (European Language Resources Association, 2020).
47. Amato, G., Falchi, F. & Vadicamo, L. Visual recognition of ancient inscriptions using convolutional neural network and Fisher vector. *J. Comput. Cult. Herit.* **9**, 1–24 (2016).
48. Avadesh, M. & Goyal, N. Optical character recognition for Sanskrit using convolution neural networks. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)* 447–452 (IEEE Computer Society, 2018).
49. Can, G., Odobez, J. M. & Gatica-Perez, D. Evaluating shape representations for Maya glyph classification. *J. Comput. Cult. Herit.* **9**, 1–26 (2016).
50. Chen, L., Lyu, B., Tomiyama, H. & Meng, L. A method of Japanese ancient text recognition by deep learning. *Proced. Comp. Sci.* **174**, 276–279 (2020).
51. Dencker, T., Klinkisch, P., Maul, S. M. & Ommer, B. Deep learning of cuneiform sign detection with weak supervision using transliteration alignment. *PLoS ONE* **15**, e0243039 (2020).
52. Hussien, R. S., Elkhidir, A. A. & Elnourani, M. G. Optical character recognition of Arabic handwritten characters using neural network. In *Proc. International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)* 456–461 (IEEE, 2015).
53. Narang, S. R., Kumar, M. & Jindal, M. K. DeepNetDevanagari: a deep learning model for Devanagari ancient character recognition. *Multimed. Tools Appl.* **80**, 20671–20686 (2021).
54. Palaniappan, S. & Adhikari, R. Deep learning the indus script. Preprint at https://arxiv.org/abs/1702.00523 (2017).
55. Suganya, T. S. & Murugavalli, S. Feature selection for an automated ancient Tamil script classification system using machine learning techniques. In *Proc. International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)* 1–6 (IEEE, 2017).
56. Burns, P. J., Brofos, J., Li, K., Chaudhuri, P. & Dexter, J. P. Profiling of intertextuality in Latin literature using word embeddings. In *Proc. 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies* 4900–4907 (Association for Computational Linguistics, 2021).
57. Pagé-Perron, E., Sukhareva, M., Khait, I. & Chiarcos, C. Machine translation and automated analysis of the Sumerian language. In *Proc. Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* 10–16 (Association for Computational Linguistics, 2017).
58. Park, C., Lee, C., Yang, Y. & Lim, H. Ancient Korean neural machine translation. *IEEE Access* **8**, 116617–116625 (2020).
59. Punia, R. N., Schenk, N., Chiarcos, C. & Pagé-Perron, É. Towards the first machine translation system for Sumerian transliterations. In *Proc. 28th International Conference on Computational Linguistics (COLING)* 3454–3460 (International Committee on Computational Linguistics, 2020).
60. Cilia, N. D. et al. An experimental comparison between deep learning and classical machine learning approaches for writer identification in Medieval documents. *J. Imaging* **6**, 89–104 (2020).
61. Reisi, E. & Mahboob Farimani, H. Authorship attribution in historical and literary texts by a deep learning classifier. *J. Appl. Intel. Syst. Inform. Sci.* **1**, 118–127 (2020).
62. Luo, J., Cao, Y. & Barzilay, R. Neural decipherment via minimum-cost flow: from Ugaritic to Linear B. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics* 3146–3155 (Association for Computational Linguistics, 2019).
63. Luo, J., Hartmann, F., Santus, E., Barzilay, R. & Cao, Y. Deciphering undersegmented ancient scripts using phonetic prior. *Trans. Assoc. Comput. Linguist.* **9**, 69–81 (2021).
64. Tupman, C., Kangin, D. & Christmas, J. Reconsidering the Roman workshop: using computer vision to analyse the making of ancient inscriptions. *Umanist. Digit.* **10**, 461–473 (2021).
65. Fetaya, E., Lifshitz, Y., Aaron, E. & Gordin, S. Restoration of fragmentary Babylonian texts using recurrent neural networks. *Proc. Natl Acad. Sci. USA* **117**, 22743–22751 (2020).
66. Bogacz, B. & Mara, H. Period classification of 3D cuneiform tablets with geometric neural networks. In *Proc. 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)* 246–251 (IEEE, 2020).
67. Dafoe, A. et al. Cooperative AI: machines must learn to find common ground. *Nature* **593**, 33–36 (2021).
68. Farzaneh, N., Williamson, C. A., Gryak, J. & Najarian, K. A hierarchical expert-guided machine learning framework for clinical decision support systems: an application to traumatic brain injury prognostication. *NPJ Digit. Med.* **4**, 78 (2021).
69. Wu, N. et al. Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Trans. Med. Imaging* **39**, 1184–1194 (2019).
70. Kim, Y., Jernite, Y., Sontag, D. & Rush, A. M. Character-aware neural language models. In *Proc. Thirtieth AAAI Conference on Artificial Intelligence* 2741–2749 (AAAI Press, 2016).
71. Ling, W., Trancoso, I., Dyer, C. & Black, A. W. Character-based neural machine translation. Preprint at https://arxiv.org/abs/1511.04586 (2015).
72. Miyamoto, Y. & Cho, K. J. Gated word-character recurrent language model. In *Proc. 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 1992–1997 (Association for Computational Linguistics, 2016).
73. Zaheer, M. et al. Advances in neural information processing systems. In *Proc. Advances in Neural Information Processes (NeurIPS)* Vol. 33 17283–17297 (Curran Associates, 2020).
74. Adhikari, A., Ram, A., Tang, R. & Lin, J. DocBERT: BERT for document classification. Preprint at https://arxiv.org/abs/1904.08398 (2019).
75. Wei, J. & Eda, K. Z. EDA: easy data augmentation techniques for boosting performance on text classification tasks. In *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* 6382–6388 (Association for Computational Linguistics, 2019).
76. Badian, E. History from "square brackets". *Z. Papyrologie Epigraphik* **79**, 59–70 (1989).
77. Bodel, J. P. Epigraphic Evidence: Ancient History from Inscriptions 52–55 (Routledge, 2001).
78. Cooley, A. The Cambridge Handbook to Latin Epigraphy 355–357 (Cambridge Univ. Press, 2012).
79. Beltrán Lloris, F. in *The Oxford Handbook of Roman Epigraphy* (eds Bruun, C. & Edmondson, J. C.) 141–143 (Oxford Univ. Press, 2015).
80. Cherry, D. Re-figuring the Roman epigraphic habit. *Anc. Hist. Bull.* **9**, 143–156 (1995).
81. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* Vol. 1 4171–4186 (Association for Computational Linguistics, 2019).
82. You, Y. et al. Large batch optimization for deep Learning: training BERT in 76 minutes. In *Proc. International Conference on Learning Representations (ICLR)* (ICLR, 2020).
83. Ghazvininejad, M., Levy, O., Liu, Y. & Zettlemoyer, L. N. Mask-predict: parallel decoding of conditional masked language models. In *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* 6112–6121 (Association for Computational Linguistics, 2019).
84. Mansimov, E., Wang, A., Welleck, S. & Cho, K. A generalized framework of sequence generation with application to undirected sequence models. Preprint at https://arxiv.org/abs/1905.12790 (2019).
85. Wang, A. & Cho, K. BERT has a mouth, and it must speak: BERT as a Markov random field language model. In *Proc. Workshop on Methods for Optimizing and Evaluating Neural Language Generation* 30–36 (Association for Computational Linguistics, 2019).
86. Schick, T. & Schütze, H. It's not just size that matters: small language models are also few-shot learners. In *Proc. 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* 2339–2352 (Association for Computational Linguistics, 2021).
87. Hornblower, S. & Matthews, E. Greek Personal Names: Their Value as Evidence (British Academy, 2000).
88. Rhodes, P. J. & Osborne, R. Greek Historical Inscriptions 404-323 BC (Oxford University Press, 2003).
89. Lewis, D. M. In *The Oxford Handbook of Ancient Greek Law* (eds Harris, E. M. & Canevaro, M.) 1–32 (Oxford Univ. Press, 2015).
90. Zelnick-Abramovitz, R. *The Concept of Manumission and the Status of Manumitted Slaves in the Ancient Greek World* (Brill, 2005).
91. Kamen, D. Sale for the purpose of freedom: slave manumission in ancient Greece. *Class. J.* **109**, 281–307 (2014).
92. Mulliez, D. Les actes d'affranchissement delphiques. *Cah. Cent. Gustave Glotz* **3**, 31–44 (1992).
93. Develin, R. Athenian Officials 684-321 BC (Cambridge Univ. Press, 1989).
94. Meiggs, R. & Lewis, D. M. A Selection of Greek Historical Inscriptions to the End of the Fifth Century B.C (Oxford Univ. Press, 1969).
95. Mattingly, H. B. The growth of Athenian imperialism. *Historia* **12**, 257–273 (1963).

96. Chambers, M. H., Galluci, R. & Spanos, P. Athens' alliance with Egesta in the year of Antiphon. *Z. Papyrologie Epigraphik* **83**, 38–57 (1990).
97. Papazarkadas, N. in *Interpreting the Athenian Empire* (eds Ma, J., Papazarkadas, N. & Parker, R.) 67–88 (Duckworth, 2009).
98. Lambert, S. D. Two inscribed documents of the Athenian empire: the Chalkis decree and the Tribute Reassessment decree. *Attic Inscr. Online Papers* **8**, 11–31 (2017).
99. Mattingly, H. B. The Athenian decree for Chalcis (IG 13.40). *Class. Q.* **52**, 377–379 (2002).
100. Lambert, S. Decrees of the council and assembly. *Attic Inscr. UK Collect.* **4**, 56–60 (2020).
101. Matthaiou, A. P. The Athenian Empire on Stone Revisited: David Lewis Lecture in Ancient History (Ellenike Epigrafike Etaireia, 2009).

Acknowledgements We acknowledge Ç. Gulçehre, P. Kohli, M. Zaheer and J. Ainslie for their scientific insights; K. Kavukcuoglu and D. Hassabis for reviewing the manuscript; J. Grayston, B. Maynard, R. Cardenas and the staff at Psycle Interactive for developing the backend of the public interface; E. Karakozoglou, A. Berger, A. Vora and D. Chou for their strategic counselling; V. Margaritis and K.-I. Naltsa for their artistic advice; and M. Tsitonas for his legal counsel. We thank R. Bagnall, L. Calvelli, A. Cooley, R. Parker and C. Roueché for their comments and insights; the contributors of PHI, LGPN, PeriodO and Pleiades for the online resources; the staff at the Acropolis museum for the permission to use the photographic reproduction of the Chalcis decree (*IG* I^3 40); and the student annotators and baseline evaluators of the Athens University of Economics and Business (MSc in Digital Methods for the Humanities) and the University of Oxford (Faculty of Classics). T.S. acknowledges that this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101026185. T.S. also thanks Harvard University's Center for Hellenic Studies for its support.

Author contributions Y.A. and T.S. are co-first authors and the order of the names is alphabetical. B.S was a major contributor to the project. M.B. contributed to the execution. J. Pavlopoulos, M.C. and I.A. contributed to the dataset analysis, the evaluations and advised the project. J. Prag and N.d.F. supervised the project.

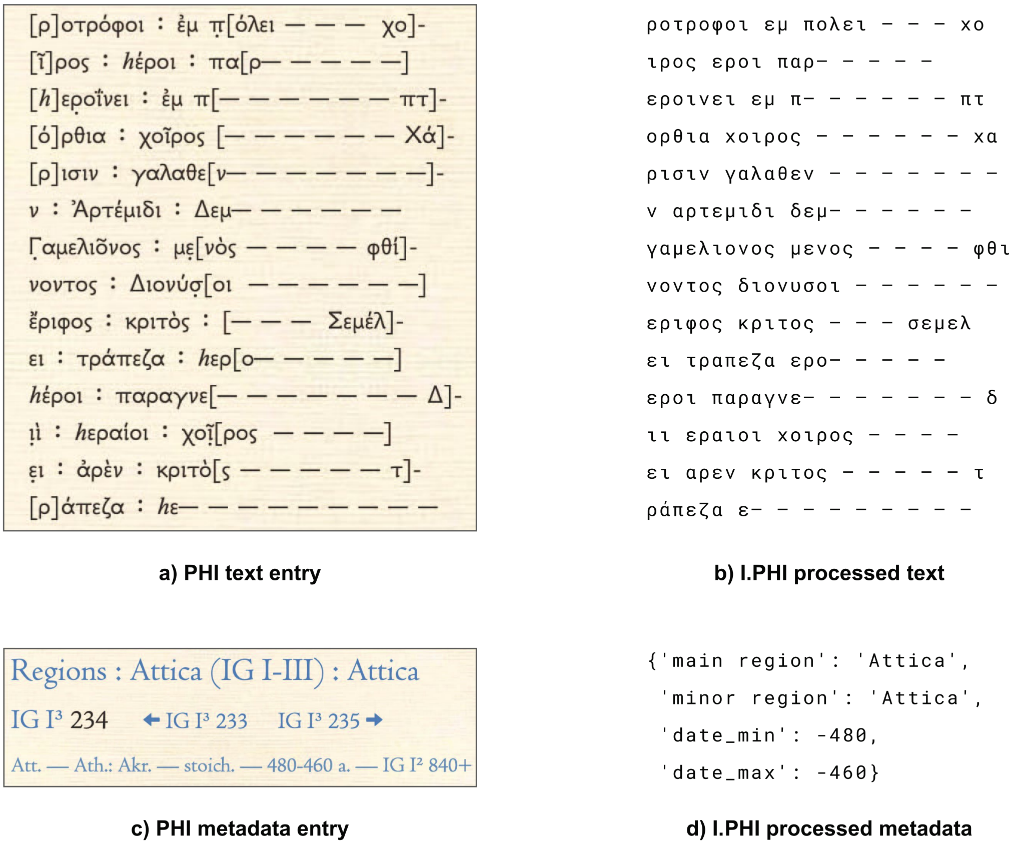
Competing interests The authors declare no competing interests.

Additional information

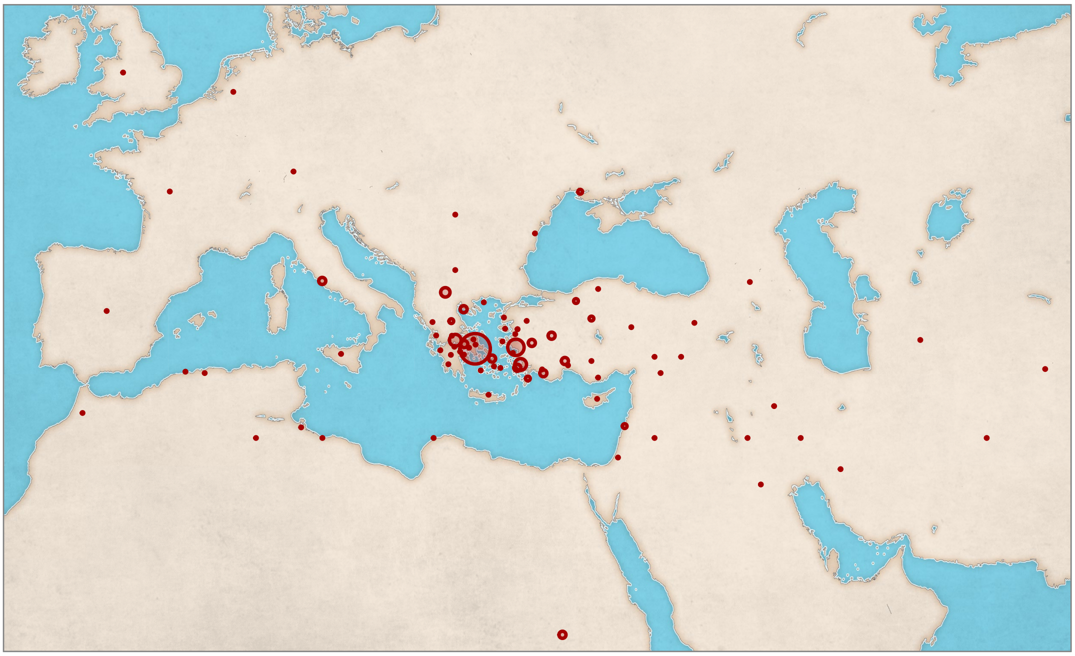
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04448-z.

Correspondence and requests for materials should be addressed to Yannis Assael or Thea Sommerschield.

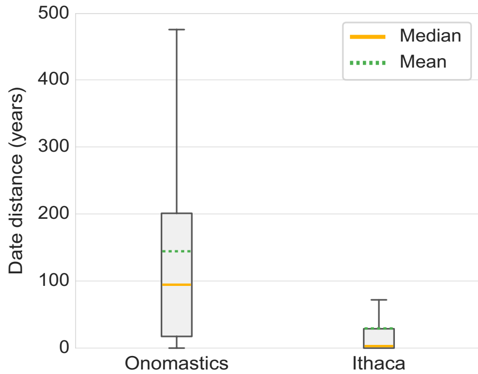
Peer review information *Nature* thanks John Bodel, Kyunghyun Cho and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Raw and processed PHI inscription, text and metadata. A fragmentary early-fifth century sacrificial calendar from the Acropolis of Athens (*IG* I[3] 234), face A, lines 10-23. In (**a**) transcription of the inscription text as it currently appears in PHI; (**b**) the same text's processed rendition in I.PHI; (**c**) the unprocessed metadata of this inscription as it currently appears in the PHI dataset; (**d**) the processed metadata rendition in I.PHI.



Extended Data Fig. 2 | Geographical distribution of Greek inscriptions in I.PHI. Each red circle represents a region across the ancient Mediterranean world (84 in total), the circle size is directly proportional to the number of inscriptions found in that region (total inscriptions in I.PHI n = 78,608).



Extended Data Fig. 3 | Comparison between Ithaca and the onomastics baseline's chronological predictions. The box plot shows the median and the mean distance between the predicted date and the ground-truth time interval, measured in years using the chronological distance metric (see Methods). In this plot, the bounds of the boxes are defined by the first and the third quartiles, and the whiskers by the minimum and maximum values. Ithaca's mean distance is 2.2x lower than that of the onomastics baseline. Ithaca's average prediction loss was 29.3 years from the ground-truth interval, while the median prediction loss was only 3 years. The onomastics baseline consists of n = 142 attributions provided by the human annotators.

Extended Data Fig. 4 | Restoration performance comparison. (**a**) The original inscription (*IG* II² 116) has 378 missing characters. (**b**) The restorations of the missing characters proposed in the authoritative edition by Rhodes - Osborne 2003 for this text[88], and which we use as ground truths in our evaluation. (**c**) Pythia's restoration shows 74 mismatches with the Rhodes-Osborne edition, while (**d**) Ithaca's shows only 45. Correct restorations are highlighted in green, incorrect ones in red.

a)

input text:	θεοι επι νικοφημο αρχοντος ------ια αθηναιων και θετταλων εις τον αει χρονον
μ:	θεοι επι νικοφημο αρχοντος ---μ--ια αθηναιων και θετταλων εις τον αει χρονον
χ:	θεοι επι νικοφημο αρχοντος ---μ-χια αθηναιων και θετταλων εις τον αει χρονον
υ:	θεοι επι νικοφημο αρχοντος -υ-μ-χια αθηναιων και θετταλων εις τον αει χρονον
σ:	θεοι επι νικοφημο αρχοντος συ-μ-χια αθηναιων και θετταλων εις τον αει χρονον
μ:	θεοι επι νικοφημο αρχοντος συμμ-χια αθηναιων και θετταλων εις τον αει χρονον
α:	θεοι επι νικοφημο αρχοντος συμμαχια αθηναιων και θετταλων εις τον αει χρονον

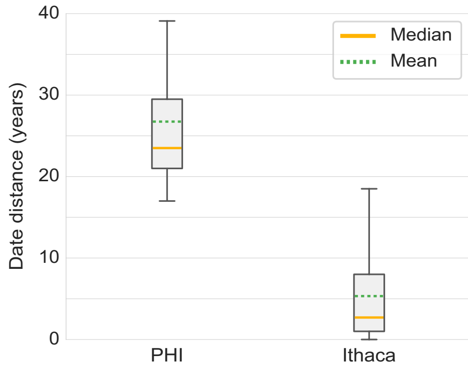
Translation: *Gods. In the archonship of Nikophemos. Alliance (συμμαχία) of the Athenians (Ἀθηναίων) and Thessalians (Θετταλῶν) for all time.*

b)

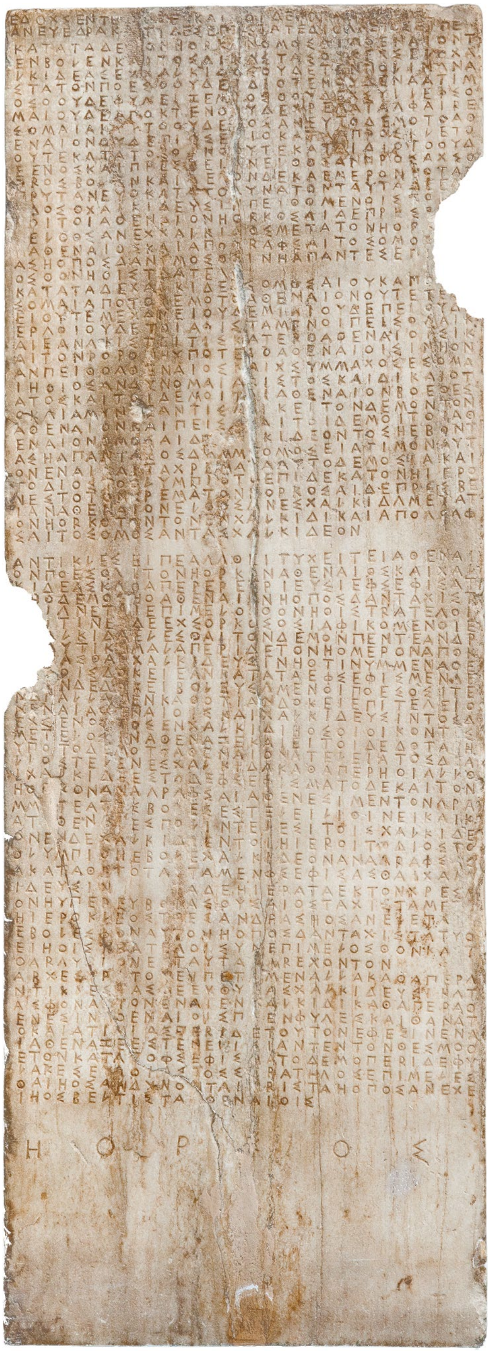
καθως επιστευσε τρυφερον τωι θεωι ταν ωναν εφ ωιτε ελευθερος ειμεν και ανεφαπτος απο παντων τον παντα χρονον ποιεουσα ο κα θελη και αποτρεχουσα οις κα θελη

Translation: *Accordingly, Trupheros entrusted (ἐπίστευσε) the sale to the god, according to which he be free (ἐλεύθερος) and untouchable by all for all time, doing (ποιέουσα) whatever he wishes and going (ἀποτρέχουσα) wherever he wishes.*

Extended Data Fig. 5 | Restoration and geographical attribution saliency maps. (**a**) The decree (*IG* II[2] 116) from the Acropolis of Athens recording an alliance between the Athenians and the Thessalian federation (360/1 BC). At each step of the restoration of the missing word "alliance" (συμμαχία), Ithaca is clearly attending to the contextually important words "Athenians" ('Αθηναίων) and "Thessalians" (Θετταλων). (**b**) The manumission inscription (*BCH* 66/67 (1942/3) 82,9) is correctly attributed to the Delphi region (left), and the generated saliency map (right) highlights words correlated to high accuracy predictions from the word statistics table.



Extended Data Fig. 6 | PHI vs. Ithaca's dating distance in years for disputed Athenian decrees. The box plot shows the median and the mean of the distribution, the bounds of the boxes are defined by the first and the third quartiles and the whiskers by the minimum and maximum values of n = 21 inscriptions. Ithaca's chronological predictions (average distance of 5 years from the modern "lower" ground truth) compared to PHI meta-data for time intervals (older estimates, average distance of 27 years from the modern ground truth). Lower distance in years is better. Exploiting the features of our full dataset, Ithaca's predictions are better and closer to modern re-evaluations compared to the original PHI ground-truth dates. The latter reflect the dates assigned by the published editions which PHI is reporting, and which almost all reflect the old three-bar sigma dating. We refer the reader to Extended Data Table 3 for detailed results.



Extended Data Fig. 7 | Chalcis decree (*IG* I[3] 40). The inscription records an oath of allegiance sworn by the city of Chalcis to Athens. It has been traditionally dated to 446/5 BC based on the 3-bar sigma criterion[28], but was more recently redated to 424/3 BC[99]. Photograph by kind concession of the Acropolis Museum. Acrop. 6509 © Acropolis Museum (photo: Socratis Mavrommatis).

Extended Data Table 1 | Dataset statistics for the size of the I.PHI corpus

Split	Inscriptions	Characters	Vocabulary	Words
Train	63,014	19,559K	209K	3,062K
Validation	7,783	2,503K	60K	391K
Test	7,811	2,415K	59K	377K

Extended Data Table 2 | Word statistics for geographical attribution

Word	Accuracy	Frequency	Prediction	Translation	Notes
ἐπίστευσε	100%	67	Delphi	entrusted	Slaves entrusted the money for their own sale to the god Apollo. The sum was then paid to the slave's master to validate the ownership transfer.
ἀποδόμενος	100%	30	Delphi	seller	The transaction took place between the master (the seller), the god (the purchaser), and the slave (the object of the transaction).
καταδουλισμῶι	100%	24	Delphi	enslavement	The god's involvement was also intended to safeguard the sale, so that freedmen would not be seized and re-enslaved.
βεβαιωτὴρ	98%	121	Delphi, Phokis	guarantor	The Delphi manumissions stipulate the conditions and price of the sale, and are certified by guarantors on behalf of the city.
ὠνὰν	93%	156	Cos, Delphi, Phokis	the sale	Over 1,350 slave manumissions are recorded by the Delphi inscriptions, offering a window onto social and demographic history.

To discover underlying patterns in Ithaca's predictions, we compute statistics to track the words that appear most frequently ("frequency") in texts whose region Ithaca predicts correctly ("accuracy"). For each word of the test set, we compute an average accuracy, and a frequency of appearance. This visualization is intended to evaluate whether the occurrence of particular words could be correlated to the model's geographical attributions.

Extended Data Table 3 | Downdating Athenian decrees with Ithaca

Subject	*IG*³ n°	PHI date (*IG*³)	New date	Ithaca prediction (mean)	PHI distance (years)	Ithaca distance (years)
Phaselis	10	469 - 450	429 - 420	399.8	50.2	20.2
Erythrae	14	circa 453/2	435/4	424.2	27.8	9.8
Egesta	11a	458/7	418/7	419.0	38.0	1.0
Sigeum	17	451/0	418/7	416.3	33.7	0.7
Coinage	1453	circa 449	425	406.5	42.5	18.5
Tribute (Kleinias)	34	448/7	425/4	424.0	23.0	0.0
Athena Nike	35	circa 448	circa 430	426.5	21.5	1.5
Eleusinian epistatai	32	circa 449 - 447	circa 432	428.4	18.6	0.0
Proxeny Delphi	27	circa 450/49	422/1	426.4	22.6	4.4
Proxeny Acheloion	19	circa 450/49	422/1	427.2	21.8	5.2
Men of Parium	18	circa 450	circa 418/7	410.9	39.1	6.1
Colophon	37	447/6	427/6	425.0	21.0	2.0
Colophon	42	circa 445 - 442	circa 425	424.3	17.7	0.7
Brea	46	circa 445	439 - 430	420.0	25.0	10.0
Eretria	39	446/5	424/3	425.0	20.0	1.0
Chalcis	40	446/5	424/3	420.3	24.7	2.7
Hestiaea	41	circa 446/5	circa 424/3	421.5	23.5	1.5
Proxeny Abydus	28	450 - 440	422/1	421.6	18.4	0.0
Miletus	21	450/49	426/5	419.5	29.5	5.5
Aegina	38	457/45	432	418.9	26.1	13.1
Hermione	31	circa 450	425/4	433.0	17.0	8.0

(all dates BCE)

List of disputed Classical Athenian decrees (including their *IG*³ edition number), their dates as listed in PHI (which follow the conventional dates proposed by Meiggs - Lewis 1969[103] and correspond to the dates in the *IG*³ editions of the decrees) based on the conventional 'three-bar-sigma' dating criterion, and their recent dating re-evaluations[28]. Ithaca's prediction mean is listed in column 5. The last two columns represent the distance (in years) of the PHI dates and Ithaca's predictions from the recent dating re-evaluations. The colour intensity reflects the distance in years, with stronger intensity reflecting a farther distance. As can be seen, Ithaca's predictions result in an average distance of 5 years, which is 22 years closer to the re-evaluated dates, compared to PHI's conventional dates.

PHI IDs of the inscriptions excluded from training: 10, 11, 14, 17, 18, 19, 27, 28, 32, 34, 37, 39, 40, 41, 42, 46, 1682; additional PHI IDs for new editions, newly discovered or published sections and doubles of the decrees: 293752, 294468, 229647, 291317, 232697, 293754, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 291118, 292366, 291960, 346490, 292187, 291318, 291321, 292189, 293756, 232710, 291322, 293327, 292194.

nature portfolio

Corresponding author(s): Yannis Assael, Thea Sommerschield

Last updated by author(s): Jan 19, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Ithaca's training and inference source code is available at https://github.com/deepmind/ithaca under Apache License 2.0, along with the trained weights, licensed under Creative Commons Attribution-ShareAlike 4.0 International (CC-BY 4.0). A public interface for historians using Ithaca for their research (i.e. restoration and attribution of Greek inscriptions, use of all visualization tools discussed in the present manuscript) is available at https://ithaca.deepmind.com. Neural networks were developed with JAX v0.2.9 (https://github.com/google/jax/), Flax v0.3.0 (https://github.com/google/flax), and Haiku v0.0.4 (https://github.com/deepmind/dm-haiku). The XLA compiler is bundled with JAX and does not have a separate version number.
Data analysis	Dataset processing and analysis used Python v3.7 (https://www.python.org/), NumPy v1.19.2 (https://github.com/numpy/numpy), SciPy v1.5.2 (https://www.scipy.org/), pandas v1.1.3 (https://github.com/pandas-dev/pandas), beautifulsoup4 v4.9.0 (https://www.crummy.com/software/BeautifulSoup/), and Google Colab (https://research.google.com/colaboratory) which is an online service and does not have a version number. Visualizations were generated using matplotlib v3.4.2 (https://matplotlib.org/), seaborn v0.11.1 (https://seaborn.pydata.org/), and GeoPandas v0.9.0 (https://geopandas.org/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Ithaca was trained on The Packard Humanities Institute's "Searchable Greek Inscriptions" public dataset, PHI, available at https://inscriptions.packhum.org/. The complete processing workflow for transforming the dataset to a machine-actionable format suitable for training Ithaca (I.PHI) is available at https://github.com/sommerschield/iphi under Apache License 2.0.

The Lexicon of Greek Person Names (LGPN) (https://www.lgpn.ox.ac.uk/) was used by annotators for the "Onomastics" baseline to track the geographical and chronological distribution of ancient names. The PeriodO gazetteer (https://client.perio.do/) was used as a reference for mapping the PHI historical time periods to the chronological range metadata of I.PHI. The Pleiades gazetteer (https://pleiades.stoa.org/) was used as a reference for mapping the PHI region names to the geographical coordinates used in the geographical attribution map visualizations.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For training the Ithaca model, no sample calculation was done, as the computational method proposed was evaluated on the full Packard Humanities Institute's "Searchable Greek Inscriptions" dataset. PHI is the largest public dataset of digitized ancient Greek inscriptions; when transforming it to a machine-actionable format, suitable for training Ithaca, it contains 78,608 inscriptions. Its large size allowed us to measure and demonstrate the effectiveness of our computational method in comparison to prior literature and the reported baselines.
Data exclusions	All inscriptions under 50 characters in length, excluding missing characters from the count, were removed; additionally, 9,441 duplicate texts were excluded.
Replication	To allow the replication of the results presented in this manuscript, Ithaca's training and inference source code is available at https://github.com/deepmind/ithaca under Apache License 2.0, along with a download link for trained weights, licensed under Creative Commons Attribution-ShareAlike 4.0 International (CC-BY 4.0). The processing workflow for transforming the Packard Humanities Institute's "Searchable Greek Inscriptions" public dataset to a machine-actionable format, suitable for training Ithaca, (I.PHI) is available at https://github.com/sommerschield/iphi, also under Apache License 2.0.
Randomization	Not applicable, we are not making a comparison between two groups.
Blinding	Not applicable, we are not making a comparison between two groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging