

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	Aeneas' source code is available at https://github.com/google-deepmind/predictingthepast under Apache License 2.0, along with the trained weights, licensed under Creative Commons Attribution-ShareAlike 4.0 International (CC-BY 4.0). Neural networks were developed with JAX v0.4.37 (https://github.com/jax-ml/jax), Flax v0.10.2 (https://github.com/google/flax). The XLA compiler is bundled with JAX and does not have a separate version number.
Data analysis	Dataset processing and analysis used Python v3.9 (https://www.python.org), NumPy v2.1.3 (https://github.com/numpy/numpy), SciPy v1.13.1 (https://www.scipy.org), pandas v2.3.3 (https://github.com/pandas-dev/pandas), beautifulsoup4 v4.12.3 (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 v3.10.0 (https://matplotlib.org), plotly v5.24.1 (https://plotly.com/python), seaborn v0.13.2 (https://seaborn.pydata.org), and GeoPandas v1.0.1 (https://geopandas.org) and CartoDB basemaps (https://github.com/CartoDB/CartoDB-basemaps) for the map tiles.

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](#)

The Latin Epigraphic Dataset (LED) was developed by processing and integrating data from the EDR, EDH, and EDCS_ETL databases, all of which are openly accessible under Creative Commons Attribution 4.0 licenses. The original data sources used in this work can be accessed directly at their respective repositories: EDR (<https://zenodo.org/records/3575495>), EDH (<https://zenodo.org/records/3575155>), and EDCS_ETL (<https://zenodo.org/records/7072337>). The processed LED dataset is publicly available online at <https://github.com/google-deepmind/predictingthepast>. The list of Roman personal names used for evaluating the Onomastics baseline was derived from the Classical Language Toolkit (CLTK) (<https://cltk.org/>), is available under the MIT License, and the processed proper names are available at <https://github.com/google-deepmind/predictingthepast>. The new curriculum co-designed as part of this research's wider knowledge exchange and impact strategy is freely available online at <https://predictingthepast.com>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Findings do not apply to only one sex or gender; sex and gender were not considered in study design; sex and gender data was not collected in this study. This is because the purpose of the study was to exclusively collect epigraphic annotations for a set of Roman inscriptions, and no participant data was recorded for this purpose.

Reporting on race, ethnicity, or other socially relevant groupings

No socially constructed or socially relevant categorisation variable(s) were used in this manuscript.

Population characteristics

No socially constructed or socially relevant categorisation variable(s) were used in this manuscript.

Recruitment

Potential participants were identified among key contributors in the field of Digital Epigraphy (with 1-4+ publications over the last 20 years) and/or among specialists in the study of Roman inscriptions. The participants were early, mid, and senior career academics in the field of Roman history, specialising in Latin epigraphy, currently employed in the Higher Education sector. In total, 41 experts were contacted via email with an invitation to participate. The 23 that accepted the call were provided with a link to the anonymous online form. Self-selection bias is possible, as participants volunteered to participate. Given the specialized nature of the study, participants' expertise mitigates potential biases affecting the study results.

Ethics oversight

Ethical approval for the AI-Historian evaluation protocol was granted by the Ethics Board of the School of Humanities, University of Nottingham. The protocol adhered to rigorous ethical standards and received a Favourable Ethical Opinion from the Faculty of Arts Research Ethics Committee at the University of Nottingham. In accordance with the approved protocol, all participants were provided with a Participant Information Sheet, a Participant Consent Form, and a GDPR Privacy Notice. A comprehensive Data Management Plan was developed, and an Awareness of Ethical Behaviour for Data Collection form was completed. Evaluation designers also completed two mandatory online courses - Research Integrity and Human Subjects Protections - offered by the University of Nottingham Researchers Academy, ensuring alignment with the highest standards of research ethics and integrity.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Behavioural & social sciences study design

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

Study description

The study was conducted in three phases:

- Phase 1: Participants were assigned a set of Latin inscriptions (five randomized texts per participant) and asked to submit their predictions for the three tasks described above using traditional methods without digital support.
- Phase 2: Participants were given another set of inscriptions (five randomized texts per participant) to restore, attribute, and date, with support from an assistive digital tool that provided its own predictions and examples.
- Phase 3: Participants completed a brief survey consisting of multiple-choice and free-form responses, assessing their experience

	using the digital aids and commenting on the usefulness of these aids to their workflow. The objective of Phase 1 was to assess the difficulty of the epigraphic tasks for domain experts using traditional methods. The objective of Phase 2 was to evaluate the extent to which digital tools could assist domain experts in completing these complex and time-consuming epigraphic tasks. The objective of Phase 3 was to gather feedback from the domain experts to evaluate their experience using digital aids for the three epigraphic tasks. The answers from each phase were scored using quantitative metrics against ground-truths, with additional qualitative feedback gathered in Phases 2 and 3.
Research sample	Potential participants were identified among key contributors in the field of Digital Epigraphy (with 1–4 or more publications over the past 20 years) and/or among specialists in the study of Roman inscriptions. To ensure inclusivity, the 23 expert researchers involved in the human evaluations reflected: 1) gender diversity (11 male, 12 female); 2) career-stage representation (early-career researchers working alongside full professors, all currently employed in the Higher Education sector); 3) age diversity (ranging from 30 to 67 years old at the time of evaluation); 4) academic specialism (all were key contributors to Roman History, particularly Latin Epigraphy, with 1 to 4 or more publications in the field over the past 20 years). In total 41 participants were contacted via email with an invitation to participate and the 23 that accepted were provided with a link to the anonymous online form.
Sampling strategy	Participants were selected through convenience sampling, specifically targeting individuals with expertise in epigraphy. Since worldwide there is a limited number of experts that could contribute to this evaluation, the sample size was defined by the number of volunteers that were willing to participate.
Data collection	The data collection took place online via an anonymous form. The responses from the participants were collected in an anonymised online form (Google Forms), and final statistics were drawn from the performance of all participants across both phases of the study. The form included the Participant Information Sheet, the GDPR Privacy Notice, and the Participant Consent Form. Participants who consented to take part proceeded to the evaluation section of the form, which was entirely anonymised. To compare the results of Phase 2 and Phase 3, participants were randomly divided into two groups, with one group conducting Phase 2 on a given inscription while the other conducted Phase 3 on the same inscription. The allocation was uniformly randomized. Each participant was assigned a minimum of five inscriptions, with the option to evaluate additional inscriptions if desired. Given the additional information provided between all Phases participants knew the Phase they were in, but didn't have visibility on the group they belonged, or answers from other participants. In total, the 23 participants evaluated 120 inscriptions, representing 60 unique texts.
Timing	The evaluations took place between 2/8/2024 and 25/09/2024. Each evaluation lasted a maximum of 2 hours.
Data exclusions	No data was excluded from the analyses.
Non-participation	No participants dropped out or denied participation. Out of the 41 participants contacted, 23 responded to our call.
Randomization	Participants were not assigned to predefined experimental groups. Instead, the study design followed a paired evaluation framework where all participants first completed tasks independently in Phase 1. In Phases 2 and 3, participants were provided with AI assistance. Each participant was presented with a set of inscriptions, and the selection of either Phase 2 or Phase 3 for each inscription was determined through uniform random sampling.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a