

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	We analyzed secondary data from the Global Preferences Survey. We used Python scripts to collect LLM-generated data via the official APIs of OpenAI, Google, and DeepSeek. Questionnaires and prompts were translated into local languages using the Google Translate API (and the GPT-4o-mini API for Dari, which Google Translate does not support).
Data analysis	We used Python (version 3.12.4) and Stata 18MP for data analysis. The following Python packages were used: NumPy (1.26.4), pandas (2.2.2), SciPy (1.16.1), Matplotlib (3.9.2), seaborn (0.13.2), Pingouin (0.5.5), statsmodels (0.14.5), and tqdm (4.66.5). In Stata, we used the norm package. All code is available in our repository (https://osf.io/pt2n4).

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](#)

We used the Global Preference Survey (GPS) dataset as the human preference benchmark. The GPS dataset is publicly available at <https://gps.econ.uni-bonn.de/home>. All LLM-generated data and the processed GPS data used in our analyses are available in our repository (<https://osf.io/pt2n4/>). GDP, population, the number of internet users, the ICT Development Index, gender-related indices, and cultural-dimension indices are publicly available. Links to access these data are provided in the Supplementary Information.

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

We used the binary gender variable contained in the GPS dataset, which serves as our benchmark. Gender information was extracted from the GPS dataset and embedded in our system prompts. We compared gender differences in preferences estimated from the LLM-generated data with those from the human survey data (i.e., the GPS dataset).

Reporting on race, ethnicity, or other socially relevant groupings

We did not use race or ethnicity data. We extracted demographic, country and region variables from the GPS dataset and embedded them into our system prompts. The definitions of all these variables are consistent with the GPS program.

Population characteristics

We used the GPS as our human benchmark and used LLMs to generate simulated preference responses for each GPS respondent, constructing a virtual GPS. The analysis sample comprises 67,201 matched respondents from 74 countries. Of these, 54.7% identified as female and 45.3% as male; the average age was 42.45 years (s.d. = 17.56 years).

Recruitment

We did not recruit any human participants for this study. All human data are drawn from the existing, publicly available Global Preference Survey.

Ethics oversight

The Ethics Committee of the Neuromanagement Laboratory, Zhejiang University, determined that this study, which involves only the analysis of existing secondary data, was exempt from ethics review.

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](https://www.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

This study used 20 LLM configurations (spanning models from OpenAI, Google, and DeepSeek) to generate simulated responses to economic preference survey items, with the Global Preference Survey (GPS) as the human benchmark. We assessed the alignment between LLM-generated data and the human benchmark data at the individual, demographic, country, and cultural levels. All data are quantitative.

Research sample

Our human data come from the Global Preference Survey (GPS) dataset, available at <https://gps.econ.uni-bonn.de/home>. The GPS elicits six experimentally validated preference measures (risk-taking, patience, positive reciprocity, negative reciprocity, altruism, and trust) from about 80,000 respondents in 76 countries. The GPS sample is representative by design. Based on the GPS dataset, we used LLMs to generate simulated preference responses for each respondent in a one-to-one manner; Supplementary Information Section A provides detailed sample information for each model and country. The GPS was chosen because it is the most comprehensive, experimentally validated, and globally representative dataset of economic preferences.

Sampling strategy

Sampling was conducted by the Global Preference Survey project. The GPS data were collected within the framework of the 2012 Gallup World Poll, and participants were selected through probability sampling designed to be nationally representative. No a priori sample size calculation was performed; we used the full available GPS sample, and the LLM-simulated sample was determined by the one-to-one matched design (one simulated response set per retained respondent per model).

Data collection	We did not collect any new human data; the human benchmark is the existing GPS dataset. The LLM-generated data were collected programmatically via the official APIs of OpenAI, Google, and DeepSeek, using Python scripts that submitted one survey item per API request, with respondent personas encoded in system prompts.
Timing	We collected GPT-4o-mini data from August 7 to August 9, 2024, Gemini 2.0 Flash and DeepSeek-V3 data from March 7 to March 19, 2025, GPT-3.5-turbo, GPT-4.1-mini, GPT-4o-mini (English) data from April 28 to May 6, 2025, GPT-4o data from September 12 to September 14, 2025, and Gemini 2.5 Flash data from September 14 to October 4, 2025. We collected GPT-5.4 and GPT-5.4-mini data (including the robustness experiments: ablations, paraphrasing, and response-constraint experiments) from March 23 to April 9, 2026, and GPT-5.4-mini (high-temperature experiment) data from May 8 to May 9, 2026.
Data exclusions	Morocco was excluded because the translation API does not support Moroccan Arabic; Cambodia was excluded because pilot data showed a high rate of non-parsable or off-scale responses. This yields 74 of the original 76 countries. Within each country, only respondents surveyed in the country's most widely sampled language were simulated. Respondents were retained in the merged analysis sample only if valid responses were obtained from all twelve LLM configurations, yielding a final sample of 67,201 respondents (approximately 84% of the original GPS sample).
Non-participation	No new human participants were recruited, so no participants dropped out or declined participation. For the LLM simulations, Morocco was not included because the translation API does not support Moroccan Arabic, and Cambodia was not included owing to a high rate of non-parsable or off-scale responses in pilot data.
Randomization	Randomization is not applicable: this is an observational audit comparing LLM-simulated responses with existing survey data, with no allocation of participants into experimental groups. Every retained GPS respondent was simulated under all LLM configurations.

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		Methods	
n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☒	☐ Animals and other organisms		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		
☒	☐ Plants		

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA