nature portfolio

Corresponding author(s): David Rand

Last updated by author(s): Oct 2, 2025

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	Qualtrics survey software (August 2024) was used to design and administer the surveys and collect data. Vegapunk (v1.0.0) was used to implement the conversation interface in Qualtrics.
Data analysis	Python (3.13.1) and R (version 4.4.1). LLMs: OpenAI GPT (4o), Deepseek (V3), Perplexity (Sonar) for AI-based fact-checking.

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

Data Availability
Due to the sensitive nature of this research—namely, the dual-use potential associated with the infrastructure and data used for political persuasion—we do not publicly share raw conversation transcripts or the bespoke LLM code. Release of these materials could facilitate misuse, such as the development or fine-tuning of

models for unethical persuasion. These materials may be made available from the authors upon reasonable request through Harvard Dataverse at https://doi.org/10.7910/DVN/DODEXZ Requests will be evaluated on an individual basis, considering the researcher's credentials, purpose, and commitment to mitigating dual-use risks.

Code Availability
Code is available at https://osf.io/vx5su/.

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	Sex and gender were not considered in the study design. Gender data has been collected and determined based on self-reporting. Consent has been obtained for sharing of individual-level data. We did not perform sex- and gender-based analyses because they were not relevant to the research questions. Gender numbers: - US election study: 1,215 females, 1,038 males, and 53 who selected another gender option - Ballot measure study: 266 females, 213 males, and 22 who selected another gender option - Canada election study: 718 females, 797 males, and 15 who selected another gender option - Polish election study: 904 females, 1,196 males, and 18 who selected another gender option
Reporting on race, ethnicity, or other socially relevant groupings	We collected standard demographic variables such as political preference/orientation (7-point scales: strongly Democratic to Strongly Republican; very liberal to very consecutive), social/economic conservatism, age, income, ethnicity, and education through self-report. We did not use these variables as proxies for other social constructs. Our studies involve randomization/experimentation, so confounding variables are unlikely to significantly affect our conclusions.
Population characteristics	See above.
Recruitment	Participants were recruited through CloudResearch Connect, Prolific, Lucid and PureSpectrum. Because our studies were related to elections and political preferences, we aimed to recruit roughly equal numbers of Democrat/liberal and Republican/conservative participants to ensure our results are generalizable and valid. Participants may have self-selected into taking part in online surveys/research and may differ systematically from the general population—for example, being more politically engaged, or having stronger opinions—this self-selection could introduce bias that limits the external validity of the findings. We conducted post hoc causal forest analysis to examine whether different subgroups might show different responses to our experimental manipulations, which helps assess the robustness of our conclusions across demographic and ideological lines (e.g., the treatment effects are actually higher among participants who are less politically engaged/interested).
Ethics oversight	MIT Committee on the Use of Humans as Experimental Subjects (protocol E-5990, E-6748, E-6757)

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	Study designs (quantitative and experimental) are detailed in the Methods section. The studies involve asking participants to engage in a multi-turn conversation with an AI/LLM chatbot about different topics/issues, and seeing if participants' preferences change after interacting with the chatbot.
Research sample	The studies recruited American/Canadian/Polish participants from CloudResearch Connect, Prolific, Lucid, or PureSpectrum. Because our studies were related to elections and political preferences, we aimed to recruit roughly equal numbers of Democrat/liberal and Republican/conservative participants to ensure our results are generalizable and valid. The samples may not be fully representative (despite our attempts to ensure representativeness) because such online participant pools tend to slightly overrepresent liberal individuals. US experiment: We preregistered a target sample of 2,400 American participants, recruiting from CloudResearch Connect and Prolific. In total, 2,457 participants started the survey, and 2,306 (1,063 Democrats, 874 Republicans, 369 Independents) completed the survey (see Tables 1 and 2). This sample (mean age = 38.13 [SD = 12.18]) included 1,215 females, 1,038 males, and 53 who selected another gender option. Sample characteristics (SDs in square brackets): AI trust = 53.83 [28.02] (0-100 scale), economic conservatism = 3.00 [1.29] (1-5 scale), social conservatism = 2.63 [1.30] (1-5 scale), belief in God = 4.09 [2.76] (0-8 scale), political interest = 60.91 [28.93] (0-100 scale) (see Fig. S1 for distributions of pre-treatment covariates). Ballot measure experiment: We preregistered a target sample of 500 Massachusetts residents. We initially pre-registered that we

would aim to recruit 250 Democrats and 250 Republicans, but given the tight timeline for subject recruitment and the limited pool of Massachusetts residents available, we decided not to implement partisanship quotas. We recruited participants from both Prolific and CloudResearch Connect. In total, 522 participants began the survey. As per our pre-registration, participants who failed to complete the pre-treatment free-text explanation of their position on the ballot measure were dropped. All who started the survey completed the pre-treatment free-text explanation, so 522 participants were assigned to treatment. Of these, another 21 dropped out during the treatment; the dropout rate did not differ significantly across conditions, $\chi2(2) = 0.49$, $p = .784$ (see SI Section S10 Fig. S30). This leaves us with a final set of 501 participants (291 Democrats, 85 Republicans, 125 Independents; mean age = 38.73 [SD = 13.74]; 266 females, 213 males, and 22 who selected another gender option). Sample characteristics (SDs in square brackets): AI trust = 4.12 [1.57] (1-7 scale), economic conservatism = 3.42 [1.75] (1-7 scale), social conservatism = 2.76 [1.62] (1-7 scale) (see Fig. S2 for distributions of pre-treatment covariates).

Canadian experiment: We preregistered a target sample of 4,000 Canadians, recruiting from four recruitment platforms (CloudResearch Connect, Lucid, Prolific, PureSpectrum) to ensure we could recruit as many participants as possible in the five days leading up to the Canadian elections on April 28, 2025. In total, 1,694 participants began the survey, and 1,530 completed the survey (850 Liberals, 490 Conservatives, 211 neither Liberal nor Conservative). This sample (mean age = 41.41 [SD = 16.67]) included 718 females, 797 males, and 15 who selected another gender option. Sample characteristics (SDs in square brackets): AI trust = 42.41 [16.67] (0-100 scale), economic conservatism = 2.89 [1.17] (1-5 scale), social conservatism = 2.51 [1.20] (1-5 scale), belief in God = 4.60 [2.71] (0-8 scale), political interest = 58.70 [28.16] (0-100 scale) (see Fig. S3 for distributions of pre-treatment covariates).

Polish experiment: We preregistered a target sample of 4,000 Poles recruited from Lucid. In total, 2,380 participants began the survey, and 2,118 completed the survey (mean age = 39.41 [SD = 20.24]; 904 females, 1,196 males, and 18 selected another gender option). Sample characteristics (SDs in square brackets): AI trust = 58.27 [27.45] (0-100 scale), economic conservatism = 2.93 [1.06] (1-5 scale), social conservatism = 2.76 [0.98] (1-5 scale), belief in God = 4.98 [2.53] (0-8 scale), political interest = 64.24 [26.10] (0-100 scale) (see Fig. S4 for distributions of pre-treatment covariates).

Sampling strategy

Sample sizes were based on previous studies of this nature (see also response to Research sample above for details), the maximum amount of money we wished to spend on each study, and the amount of time we had leading up to each election (given the research questions, we had to end data collection before election day). We had no strong basis for an a prior estimated effect size and such power analyses are often uninformative (http://datacolada.org/4).

Data collection

We used Qualtrics survey software to collect data from online recruitment platforms. The researcher was not present with the participant when data collection happened (so the researcher was blinded to the experimental condition). Given data collection happened online/remotely, we cannot determine whether someone else was with the participant.

Timing

The US experiment was conducted from Aug 22 to Nov 2, 2024. The ballot measure experiment was conducted from Oct 30 to Nov 5, 2024. The Canadian experiment was conducted from April 23 to April 28, 2025. The Polish experiment was conducted from May 7 to May 16, 2025.

Data exclusions

We excluded participants for inattentiveness before the study using an open-ended text response. In addition, in the ballot measure study, participants who failed to complete the pre-treatment free-text explanation of their position on the ballot measure were excluded.

US experiment: 2,457 participants started the survey, and 2,306 (1,063 Democrats, 874 Republicans, 369 Independents) completed the survey (see Tables 1 and 2).

Ballot measure experiment: In total, 522 participants began the survey. As per our pre-registration, participants who failed to complete the pre-treatment free-text explanation of their position on the ballot measure were dropped. All who started the survey completed the pre-treatment free-text explanation, so 522 participants were assigned to treatment. Of these, another 21 dropped out during the treatment; the dropout rate did not differ significantly across conditions.

Canadian experiment: 1,694 participants began the survey, and 1,530 completed the survey (850 Liberals, 490 Conservatives, 211 neither Liberal nor Conservative).

Polish experiment: 2,380 participants began the survey, and 2,118 completed the survey.

Non-participation

151 of 2,457 participants dropped out in the US experiment; 21 of 522 participants dropped out in the ballot measure study; 164 of 1,694 participants dropped out in the Canadian experiment; 262 of 2,380 participants dropped out in the Polish experiment. We reported analyses in the SI showing that there was no evidence for differential attrition/dropout across experimental conditions. Because of the online nature of the studies, we are unable to collect data on why the participants dropped out.

Randomization

Randomization was implemented using Qualtrics survey software.

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

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☐	☒	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☐	☒	Any other significant area

Other impacts: Democratic processes and election integrity

Hazards: LLM chatbots can influence beliefs and voting behavior, risking democratic integrity.

For examples of agents subject to oversight, see the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern.

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☐	☒	Any other potentially harmful combination of experiments and agents

Other combinations: Data/pipelines from our experiments can be used to train/improve/fine-tune AI models to improve their persuasive capabilities.

Precautions and benefits

Biosecurity precautions: *Describe the precautions that were taken during the design and conduct of this research, or will be required in the communication and application of the research, to minimise biosecurity risks. These may include bio-containment facilities, changes to the study design/ methodology or redaction of details from the manuscript.*

Biosecurity oversight: *Describe any evaluations and oversight of biosecurity risks of this work that you have received from people or organizations outside of your immediate team.*

Benefits: *Describe the benefits that application or use of this work could bring, including benefits that may mitigate risks to public health, national security, or the health of crops, livestock or the environment.*

Communication benefits: *Describe whether the benefits of communicating this information outweigh the risks, and if so, how.*

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*