nature portfolio

Corresponding author(s): DAPR NCOMMS-25-36258

Last updated by author(s): Jun 12, 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	No software was used. Secondary data used: Raleigh, C., Linke, rew, Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. Journal of Peace Research, 47(5), 651–660. https://doi.org/10.1177/0022343310378914 Dataset is currently provided as Supplementary Data to the manuscript.
Data analysis	All analyses were conducted using Python 3.12.5. All necessary packages and versions are provided through a virtual environment and "requirements.txt" file currently provided in the supplementary materials of the manuscript. For the analysis, GPT 4 and GPT 4-mini were accessed through OpenAI API keys. To run the code, users have to insert their own API key at the highlighted positions in the code. Even without API key, figures etc. are reproducible. The user can skip the API-related code snippets.

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 dataset is publicly available and stems from "acleddata.com". All data are currently provided as a supplement to the submitted manuscript. If the user wishes to download the data themselves, they can do so through signing up for an ACLED API key. More information on that can be found on "apidocs.acleddata.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	The study does not include any analysis or information of sex or gender of the participants. The study uses secondary data provided by the Armed Conflict Location and Event Data Project (see above). No socioeconomic information is included.
Reporting on race, ethnicity, or other socially relevant groupings	The data collection process is based on filtering criteria: protests events occurring between November 1, 2023, and March 31, 2024, protests took place in Europe, involving either farmers or organizations associated with them. Hence socially relevant groupings involve farmers or organizations associated with them because the analysis studies farmer protests in Europe. No statistical analysis was conducted.
Population characteristics	See above
Recruitment	All the data was sourced from Armed Conflict Location and Event Data Project (see above). Several biases might be relevant for the analysis and are discussed in detail in the discussion section of the manuscript: Conflict and protest data, like that used by ACLED, often comes with inherent challenges related to reporting bias, which may stem from systematic inclinations, prejudices, and the directionality of information, influencing how these events are portrayed, how they are selected, and which aspects are emphasized (Miller et al., 2022; Weidmann, 2016). Although ACLED has strong quality control mechanisms (ACLED, 2023) that help reduce bias and maintain data reliability (see Online methods), it is impossible to completely eliminate bias from the process. Additionally, the LLM classifier cannot be tested against a ground truth, as the true nature of many events is inherently open to interpretation. This limitation, however, is not unique to machine-based classification; human annotators face the same challenge when classifying events, as their assessments are similarly influenced by context, perspective, and available information (Aroyo & Welty, 2015). To address this concern, we present evidence of agreement among human and machine annotators in the Supplementary Information.
Ethics oversight	The study protocol was reviewed, and it was determined that formal ethical approval was not required as the research did not involve sensitive data. No ethical concerns were identified. This determination was made in accordance with the institutional policies of ETH Zurich.

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	This study employed the analytical capabilities of large language models (LLMs) to examine publicly available textual narratives concerning large-scale farmer protest events across Europe during 2023–2024. The research involved computational analysis of existing textual data without the collection of personal, sensitive, or identifiable information.
Research sample	The study uses an existing dataset.
Sampling strategy	The data used in this study comes from the Armed Conflict Location & Event Data Project (ACLED), a comprehensive database that specializes in collecting and analyzing information on political violence, protests, and conflicts in various regions of the world. ACLED's data collection follows a methodology that draws from a wide range of sources, including international, national, and local news agencies, nongovernmental organizations (NGOs), government reports, social media, and direct reports from eyewitnesses and

regional partners. This multi-source approach ensures comprehensive coverage, especially in underreported or remote areas, and provides a more complete picture of global unrest.

By using this comprehensive dataset, we are confident that we have achieved an extensive and as-complete-as-possible coverage of the European farmer protests 2023-2024 —while acknowledging that absolute completeness cannot be guaranteed.

Data collection

We filtered the ACLED dataset by the following criteria:

-Events between November 1, 2023, and March 31, 2024.
-Protests took place in Europe
-Either farmers or organizations associated with them were involved

Timing

The data were downloaded on 2024-09-10.

Data exclusions

We filtered the original ACLED dataset by the following criteria:

-Events between November 1, 2023, and March 31, 2024.
-Protests took place in Europe
-Either farmers or organizations associated with them were involved

339 recorded events did not contain an event description and were not used for the analysis of protest reasons.

The data filtering process is outlined in the online methods section.

Non-participation

No participants dropped out.

Randomization

Data collection was conducted by ACLED (see above).

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

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.