



OPEN Harnessing deep learning to monitor people's perceptions towards climate change on social media

Ana Sofia Cardoso^{1,2,3✉}, Catarina da Silva^{1,2,3}, Andrea Soriano-Redondo^{4,5}, Ivan Jarić^{6,7}, Susana Batel⁸, João Andrade Santos⁹, Alípio Jorge^{10,11} & Ana Sofia Vaz^{1,3,12}

Social media has become a popular stage for people's views over climate change. Monitoring how climate change is perceived on social media is relevant for informed decision-making. This work advances the way social media users' perceptions and reactions towards climate change can be understood over time, by implementing a scalable methodological framework grounded on natural language processing. The framework was tested in over 1771 thousand X/Twitter posts of Spanish, Portuguese, and English discourses from Southwestern Europe. The employed models were successful (i.e., > 84% success rate) in detecting relevant climate change posts. The methodology detected specific climate phenomena in users' discourse, coinciding with the occurrence of major climatic events in the test area (e.g., wildfires, storms). The classification of sentiments, emotions, and irony was also efficient, with evaluation metrics ranging from 71 to 92%. Most users' reactions were neutral (> 35%) or negative (> 39%), mostly associated to sentiments of anger and sadness over climate impacts. Almost a quarter of posts showed ironic content, reflecting the common use of irony in social media communication. Our exploratory study holds potential to support climate decisions based on deep learning tools from monitoring people's perceptions towards climate issues in the online space.

Keywords Sentiment analysis, Emotion analysis, Environmental monitoring, Irony detection, Natural Language processing, Twitter

Climate change brings a wide spectrum of consequences for human subsistence and wellbeing, ranging from phenological changes in vegetation to shifts in wildlife distributions, or changes in both abiotic and biotic stresses, affecting food and water production systems, and challenging human health and socioeconomic revenues¹. The recognition that climate change represents a serious threat has been taking place over the global political agenda and human society, and the prevention and mitigation of its impacts have become among the biggest societal priorities^{2,3}.

Dealing with climate change represents a major challenge for both the natural and social realms, as climate related issues are shaped by a diversity of views, narratives and discourses intrinsically embedded in individual, institutional and cultural contexts^{1,4}. Exploring how climate change is perceived by the public is a multifaceted

¹CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal. ²Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto 4099-002, Portugal. ³BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão 4485-661, Portugal. ⁴Departamento de Anatomía, Biología Celular y Zoología, Facultad de Ciencias, Universidad de Extremadura, Badajoz, Spain. ⁵Ecología en el Antropoceno, Facultad de Ciencias, Unidad Asociada CSIC-UEx, Universidad de Extremadura, Badajoz, Spain. ⁶Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Gif-sur-Yvette, France. ⁷Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, Ceske Budejovice, Czech Republic. ⁸Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, Lisbon, Portugal. ⁹Centre for the Research and Technology of Agroenvironmental and Biological Sciences, CITAB, Institute for Innovation, Capacity Building and Sustainability of Agri-food Production, Inov4Agro, Universidade de Trás-os-montes e Alto Douro, UTAD, Vila Real 500-801, Portugal. ¹⁰LIAAD, INESC TEC, Rua Dr. Roberto Frias, Porto, Porto 4200-465, Portugal. ¹¹Faculty of Sciences, University of Porto, Rua do Campo Alegre, Porto, Porto 4169-007, Portugal. ¹²NBI, Natural Business Intelligence, Régia Douro Park, Andraes, Vila Real 5000-033, Portugal. ✉email: sofia.cardoso@cibio.up.pt

task that requires the consideration of cognitive, emotional, and contextual factors. Previous studies suggest that perceptions of climate change are not monolithic but vary widely across different populations^{5,6}.

Perceptions over climate change are associated to a range of sentiments and emotions, from concern and anxiety to scepticism and indifference, often associated also to a diversity of factors, such as media representations, socio-political contexts, and personal experiences with climate-related events⁷. For instance, people who have directly experienced extreme weather events may have heightened concern and urgency regarding climate change, whereas others might be more influenced by political ideology or the framing of information in the media^{8,9}.

The emergence of climate change narratives on social media platforms adds a new layer of complexity into the way these issues are communicated and perceived¹⁰. Social media offers a stage for communicating about climate change issues, a stage that is characterised by the immediacy and brevity of posts, the use of multimedia elements, or the fast interactive nature of the platforms. These characteristics enable the rapid dissemination of information, foster real-time discussions, and allow for the use of visual aids like images and videos to enhance the impact of messages¹¹.

Nonetheless, the decentralized and user-driven nature of social media implies that information may be fragmented and vary widely in accuracy and tone. Recent studies have highlighted that user-generated content, particularly in the scope of climate change, often reflects personal experiences and opinions, which can lead to a diverse range of narratives, from highly scientific and data-driven posts to sarcastic and emotive accounts^{5,12}. Moreover, social media algorithms tend to create echo chambers, where users are exposed primarily to content that reinforces their existing beliefs. This can amplify polarization and create distinct online communities with differing views on climate change¹³.

Although it is still not clear whether the overall interest in climate change and other environmental issues is increasing or not, the use of online resources to automatically understand people's awareness of and ideas about climate change is increasingly recognized as an important area of research^{14,15}. Therefore, assessing and monitoring social media data on climate change is invaluable for obtaining (near) real-time insights into public sentiment and awareness, enabling researchers and policymakers to gauge the level of concern and engagement within different communities¹⁶.

Dealing with the massive amount of social media data produced by an increasing number of social media users is technically (and ethically) challenging, especially the extraction of meaningful information for research¹⁷. Recent technological advances in artificial intelligence algorithms, such as those from computer vision and natural language processing (NLP), have empowered social media research to address several issues regarding public's perceptions of climate change^{18,19}, as well as of specific phenomena, such as the severity of floods²⁰, or responses to climate disasters²¹.

Adopting methods for automated analyses of social media content could leverage the potential of digital technology to support decision-making through the social monitoring of climate change in the online space. For instance, a research study²² developed a deep learning-based sentiment analysis system to gauge public sentiment on climate change from social media data, providing insights that inform smart city governance and climate policies. Similarly, an analogous study²³ employed sentiment and emotion analysis using deep learning to assess public reactions to climate change and energy issues in the United Kingdom and Spain, providing valuable insights for policymakers to tailor their communication strategies and interventions. Despite the opportunities of online digital data and artificial intelligence for the monitoring of people's perceptions, their combined application in the field of climate change is still in its infancy.

Even though research on climate change perceptions is growing, significant gaps remain in understanding public sentiments and emotions across different social media platforms. While previous studies have explored climate change discourse on social media, they often focus on specific events or regions, and many rely on traditional sentiment analysis techniques that may not fully capture the complexity of human emotions and perceptions^{20,24}. In contrast, our methodology not only assesses emotions and expressions over time, but also identifies the most mentioned climate change phenomena and impacts, offering a more comprehensive understanding of public discourse on climate change. Furthermore, while deep learning techniques have been applied to various domains, their use in automated, large-scale monitoring of climate change perceptions on social media is still in its early stages^{19,22}.

In this study, we aim to fill these gaps by testing the ability of using pre-trained and open-source deep learning models to support the social monitoring of climate change from textual content shared on social media platforms. Using Southwestern Europe as study area, we implement a scalable methodological framework grounded on NLP and textual content from the social media platform Twitter (currently known as X). By implementing the proposed methodology, we: (1) identify the most mentioned climate change phenomena and impacts by social media users and (2) assess the users' perceptions, sentiments, and emotions towards climate change through time. We further discuss the major caveats and opportunities for leveraging scalable methodologies for real-time social monitoring, to support climate change research and informed decision making.

Results

Relevant climate change content in social media posts

Overall, the natural language processing (NLP) models employed under our proposed methodology showed high performances when detecting relevant posts containing climate change related content over the social media dataset. The models showed f1-score values above 87% for Portuguese, Spanish and English posts (Table 1). The best performance was obtained for English posts (BERT-base model) in terms of precision (94% and 95%), recall (94% and 96%) and f1-score (95%). The Spanish model (BETO) also showed high precision (92% and 93%), recall (91% and 94%) and f1-score (92% and 93%). Conversely, the Portuguese model (BERTimbau) showed the lowest, yet satisfactory, performance with 89% and 85% precision, 84% and 90% recall and f1-score of 87%.

	Precision	Recall	F1-score
BERT-base model (English posts)			
Irrelevant posts	0.95	0.94	0.95
Relevant posts	0.94	0.96	0.95
BERTimbau model (Portuguese posts)			
Irrelevant posts	0.89	0.84	0.87
Relevant posts	0.85	0.90	0.87
BETO model (Spanish posts)			
Irrelevant posts	0.92	0.94	0.93
Relevant posts	0.93	0.91	0.92

Table 1. Classification report of the BERT models trained with a sample of annotated Twitter posts. Bold values highlight the best performance results.

Climate change phenomena in social media posts through time

The time series and frequency analyses integrated in our methodology detected specific climate phenomena in social media users’ discourse, coinciding with the occurrence of major climatic events in the test area (Fig. 1). Peaks of social media posts were found to be significantly high at the time major meteorological and climate-related phenomena occurred in the test area, including large scale wildfires (August 2012; June and October 2017; August 2019; July and August 2022), and powerful storms (September and December 2019) (Fig. 1, see Fig. S5 for details).

Sentiments of social media users towards climate change

The open-source NLP models implemented in our approach exhibited good performances when classifying and detecting sentiments, emotions, and irony in social media posts, with performance values from 71 to 92% (Table 2). Most users’ sentiments across the analysed period were found to be neutral (> 35% of all relevant posts) or negative (> 39%). Negative posts included mostly content on climate impacts pertaining to pests, pollution, wildfires, or droughts. Only a minor portion of the analysed posts showed positive sentiments, mostly associated with terms pertaining to sustainability or climate action (> 13%; Figs. 2a and 3).

Emotions and irony towards climate change in social media content

The emotion analysis suggested a predominance of posts expressing neutral emotions; from 61% to up to 74%, depending on the analysed year. Anger, surprise and joy were the most prevalent emotions in social media posts (Fig. 2b). An increasing frequency of posts with content associated with anger was also noticed through time, and particularly over the last 4 years covered by the analysis. Ironic content was detected in ca. quarter of the dataset (Fig. 2c), showing a steady prevalence over the last decade.

Discussion

The results from our exploratory study already suggest potential to support climate decisions based on deep learning tools from monitoring people’s perceptions towards climate issues in the online space. The application of the NLP models to the test area showed good performance in detecting relevant social media content for monitoring climate change phenomena, as well as for tracking people’s perceptions towards climate-related impacts.

Our time series and text mining analyses showed prevalence of posts particularly referring to storms, wildfires, and heatwaves, converging with times when such phenomena mostly took place in the test area (Fig. S2)^{25,26}. Curiously, other climatic hazards, such as droughts and extreme low temperature, which have also been recurrent in the test area²⁷ were not included among the most prevalent phenomena in the posts. In the absence of further information about these patterns, we can suggest two complementary hypotheses to explain these observations. First, people tend to be more aware about phenomena with more direct and immediate socio-economic impacts²⁸. Second, weather episodes, and their effects, are likely to be more recognized by individuals who have directly experienced them²⁹.

This may possibly be the case in the test area, in which previous studies³⁰ have already suggested that the social memory of public officers on the effects of wildfires, storms and heatwaves may be more prominent than droughts, and that a direct link between droughts and wildfire occurrence might not be obvious. When it comes to social media users, results observed are probably associated with specific climate change events that occurred during the temporal periods of the most prominent posting peaks. For instance, during June 2017, a series of four severe wildfires took place in Portugal, which resulted in around 66 deaths and 204 injured people³¹. Similarly, in October 2019, more than 500 large wildfires became simultaneously active in both Portugal and Spain, with the passage of the tropical storm Ophelia³². During July and August 2022, associated with a major heatwave, over 57,000 hectares were destroyed by wildfires in Portugal and Spain, causing at least 135 injuries, as well as the evacuation of around 2300 people³³. Other extreme weather events, such as tropical storms Elsa, Fabien, and Gordon, also affected the test area in December 2019 and August 2012^{34,35}.

Although social memory and impact perception may drive peaks of communications on social media, it is also acknowledged that major or global public events can lead to instant and prominent reactions on social media³⁶. From our analyses, we can note particular social media activity on climate change content in 2017 and

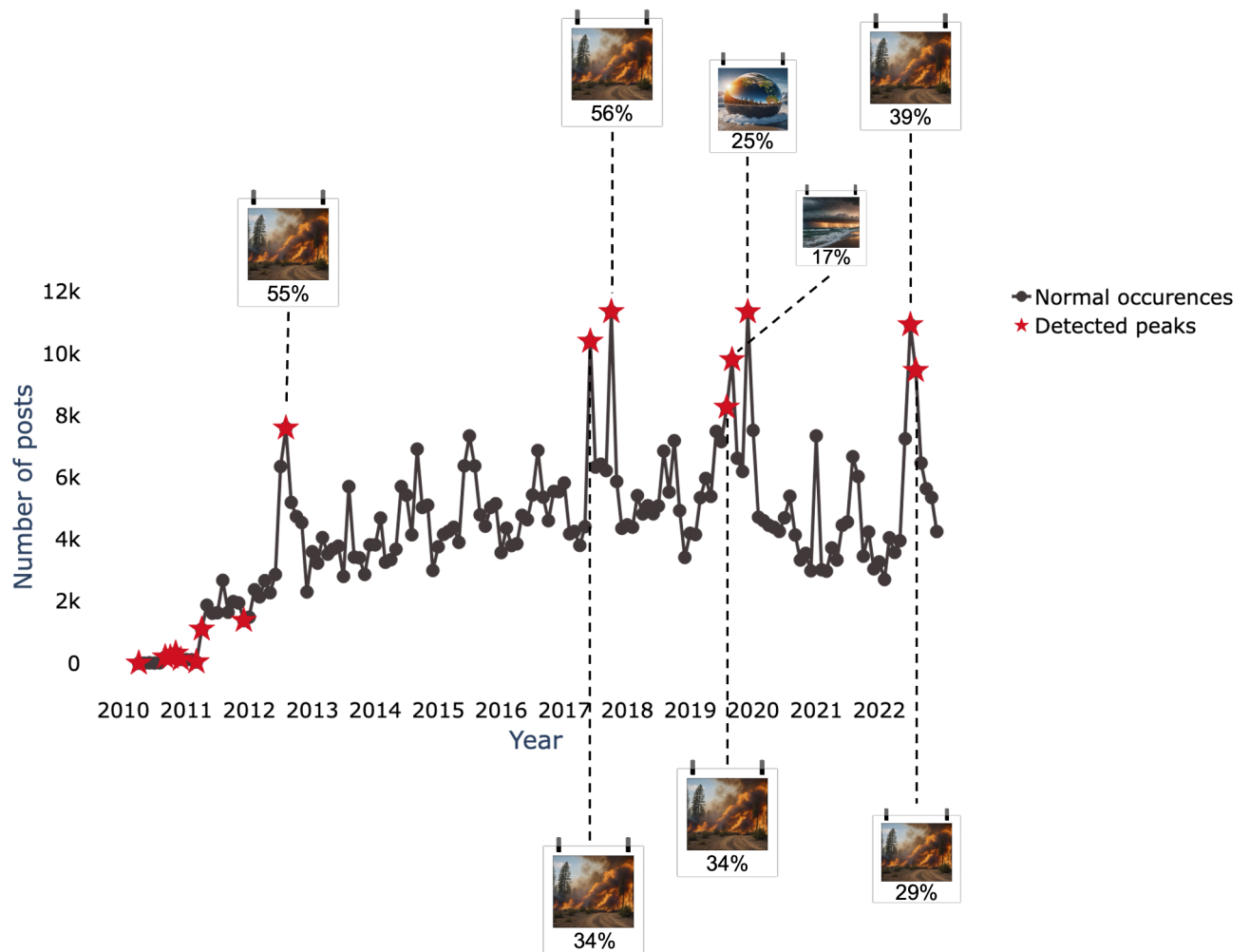


Fig. 1. Time series with the frequency of posts between 2010 and 2022. Red stars indicate the peaks identified in the anomaly detection analysis. Peaks of social media posts were significantly high during major meteorological and climate-related phenomena in the test area, including large-scale wildfires (August 2012; June and October 2017; August 2019; July and August 2022) and impactful storms (September and December 2019). The images were generated using the prompt “Wildfires”, “Storms” and “Global warming” by Gencraft, Image 2.0, 2024 (<https://gencraft.com/>).

2019, when the Conference of the Parties (COP) on climate change took place in Bonn (Germany, COP23) and Madrid (Spain, COP25), respectively. COPs, mostly the ones that address climate change, besides being extremely relevant events at a global scale, are always very popular topics on network platforms online. Therefore, there is usually an increase in the number of posts before, during and after the end of these events. The fact that COP25 took place in Madrid probably drove an additional increase in the number of posts (December 2019).

When we applied the sentiment analysis models, most posts revealed a predominance of negative or neutral sentiments over time. In the case of negative sentiments, posts were particularly related to extreme weather events and natural hazards, whereas in the case of neutral posts they were more closely associated with terms related to climate action. This observation is in line with previous studies that also found general negativity associated with climate change online discourses, especially when content pertains to natural disasters with more immediate damages and threats^{19,37}. Conversely, general climate change terms, such as global warming and the greenhouse gas effects, tend to be seen as rather abstract, gradual, and often distant from the immediate concerns of daily livelihoods³⁸. Likewise, neutral posts are usually popular on social media networks when users report on news, facts, or event announcements in a more objective content.

Over the last years, we notice a shift in the nature of posts, with an increasing emergence of content suggestive of anger, joy, or surprise. This pattern may be due to an overall discontent or an increase in users' interest towards climate change impacts and policies, which might have been influenced by specific events that occurred during those years. Specifically, the change in emotions could be fostered by an increase in the occurrence of extreme weather events or environmental disasters, the implementation of controversial climate-related policies/actions by governments, the occurrence of high-profile global events or climate change campaigns, among others, which may evoke strong emotional reactions, leading to a rise in people's frustration or concern^{39,40}. For irony

	Accuracy	Precision	Recall	F1-score
Sentiment analysis				
Portuguese	92.00	85.06	89.81	87.12
Spanish	83.81	80.98	88.12	82.54
English	86.00	82.02	90.73	85.10
Emotion analysis				
Portuguese	86.70	81.32	79.74	78.44
Spanish	91.70	86.54	79.12	77.63
English	90.80	86.50	84.59	84.70
Irony detection				
Portuguese	84.00	74.57	87.27	77.62
Spanish	84.40	72.26	88.14	75.81
English	85.80	70.87	90.39	75.07

Table 2. Evaluation metrics computed to test the robustness of the sentiment, emotion, and irony detection models in Portuguese, Spanish and english. Bold values highlight the best performance results per task. Bold values highlight the best performance results.

detection, in turn, the results were more consistent over the years, with most posts expressing unironic content. These findings align with those from Becker & Anderson⁴¹, who found that, while humorous and ironic content can attract attention, is less prevalent in serious discussions about climate change. Nevertheless, it is widely recognized that irony is part of communication channels of many cultures, and so is in both Portugal and Spain. For instance, in Portuguese and Spanish cultures, irony is often used by social media users as a tool for expressing disagreement or critique without direct confrontation on societal issues, particularly in the context of media, or political public debates^{42,43}. The platform’s own configuration—its character limits, real-time flow, and emphasis on quick engagement—may shape how users interact and express themselves. This could be one of the reasons behind observed results, as this setup encourages succinct and sharp communication, pushing users to condense their thoughts into brief, yet impactful statements⁴⁴. However, in the light of the limited information from our study, and the lack of wider studies on the issue, this leaves us no concrete insights to further understand the ironic patterns in the studied test area.

Lastly, when analysing the individual sentiments, we noticed overall negativity associated with most climate change phenomena, except for posts mentioning environmental issues, decarbonization, weather phenomena, climate emergency, climate action, eutrophication, landslides, and climate goals, which revealed more neutral sentiments. For instance, sustainability, climate action, climate goals and decarbonization are topics that arouse, in general, positive perceptions, since many users are passionate about the planet, often advocating for positive change and supporting green initiatives, to encourage people to participate in collective efforts for a better future^{45,46}.

While the preliminary results of this study are promising, they show some limitations and caveats. First, social media users may not represent the full range of society, as individuals who do not use social media are excluded⁴⁷. This introduces representation bias, where certain demographic groups may be underrepresented or entirely absent, potentially biasing the findings. Second, social media platforms often facilitate echo chambers, where users primarily interact with others who share similar perspectives¹³. This phenomenon may lead to an overrepresentation of certain perspectives or sentiments while suppressing others, thus influencing the generalization of the results. A potential step forward to address these limitations could be expanding the research to other social media platforms, which may offer distinct user dynamics and engagement patterns, leading to a more comprehensive understanding of public perceptions on climate change. Third, despite the great performance of the employed models, no model is flawless. Errors may persist in automated analyses, potentially excluding relevant posts related to climate change or misinterpreting sentiments, emotions, or ironic content⁴⁸. This is especially true for the detection of irony and sarcasm, which is linguistically and culturally dependent, being much more complex to assess. Moreover, labelling sentiments and emotions is usually a significantly subjective process, with different individuals having distinct interpretations⁴⁹. The study’s reliance on these labels may introduce subjectivity, impacting the reliability of the results.

Despite its limitations, we are confident that our insights hold opportunities to support informed decisions and research in the field of climate change^{50,51}. Overall, our study extends the scope of existing literature by showcasing the benefits of using advanced NLP tools to detect complex dimensions of social media climate discourse, such as emotions and irony. This granular approach provides a more detailed understanding of the public’s emotional landscape, revealing patterns such as the association of negative emotions with extreme weather events and neutral tones with news and announcements^{52,53}. Such findings can inform communication strategies to better engage the public, tailor responses to climate events, and enhance the effectiveness of climate change campaigns in fostering action and resilience. Likewise, understanding how sarcasm, humour, or irony are embedded in climate-related discussions can mitigate potential misinterpretations and reveal opportunities for more effective engagement with audiences who use irony as a form of critique or expression⁵⁴.

Temporal pattern analysis of social media content can reveal trends in public interest and awareness regarding climate change, aiding the management of misinformation, prevention of upcoming crises, and the strategic

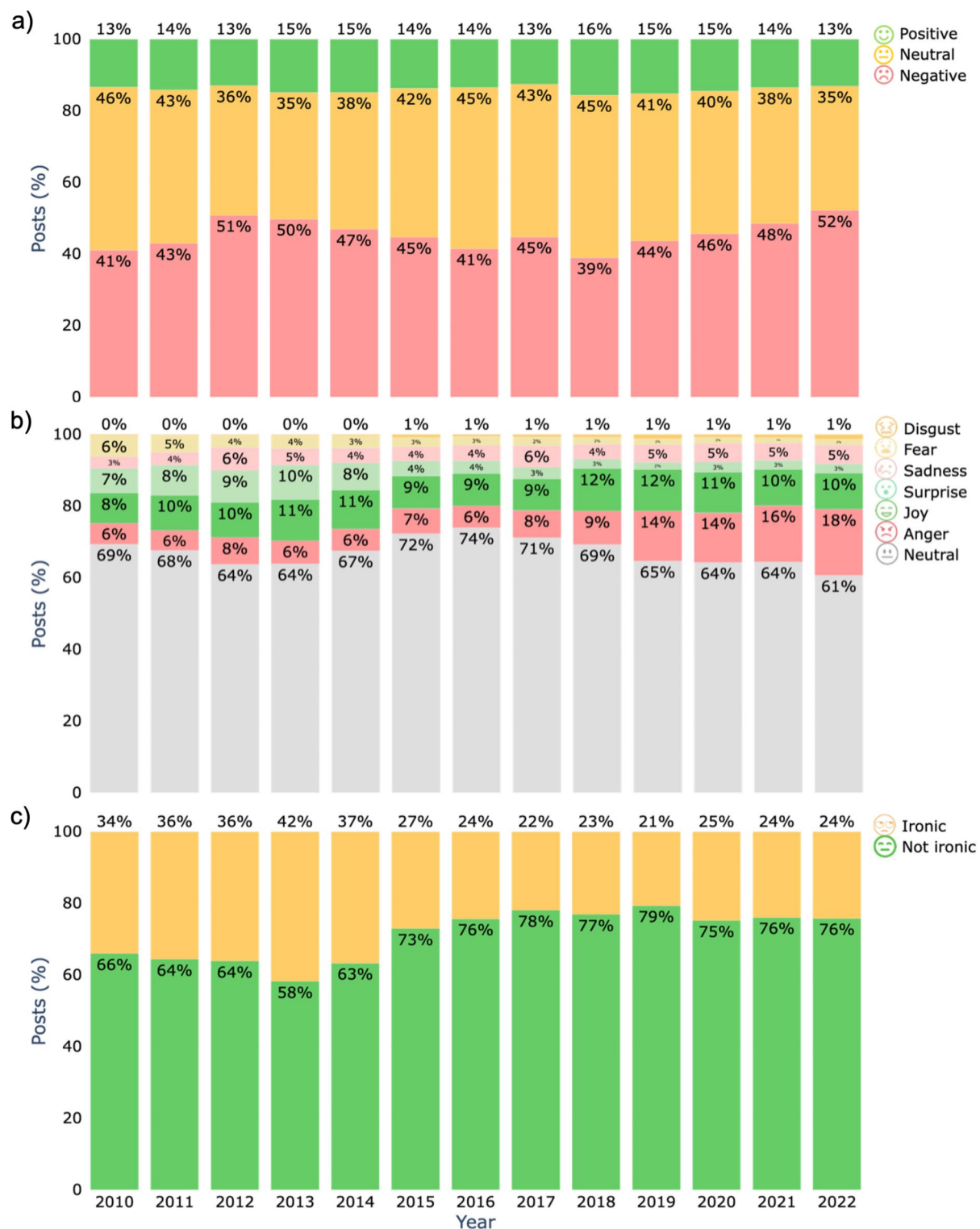


Fig. 2. Relative frequency of (a) positive, neutral, and negative posts, (b) posts expressing anger, joy, sadness, fear, surprise, disgust and neutral emotions, and (c) ironic and not ironic posts, between 2010 and 2022. Most users' sentiments across the analysed period were neutral (> 35% of all relevant posts) or negative (> 39%). Positive sentiments accounted for a minor portion of posts (> 13%). Emotion analysis showed a predominance of neutral emotions (61–74%, depending on the year), with anger, surprise, and joy being the most frequently detected emotions. An increasing trend of anger-related posts was observed, particularly over the last four years. About a quarter of the dataset contained ironic content, showing a steady prevalence over the last decade.

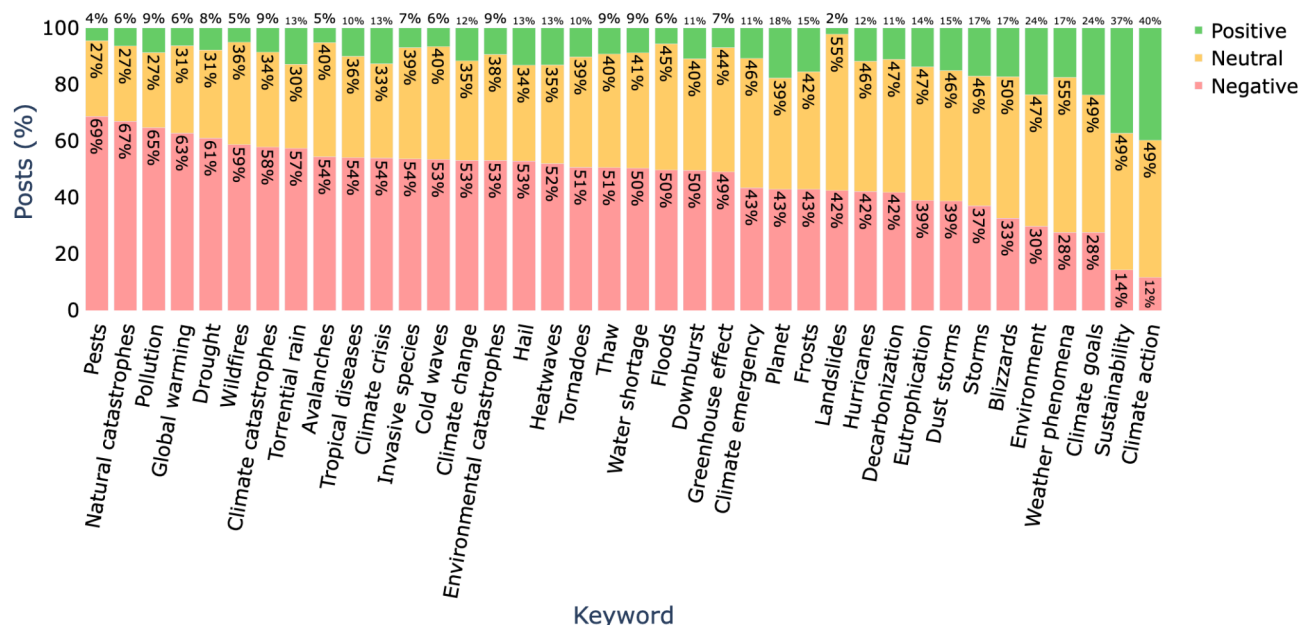


Fig. 3. Relative frequency of positive, neutral, and negative posts for each phenomenon. Negative posts predominantly included content related to climate impacts such as pests, pollution, wildfires, or droughts. In contrast, only a minor portion of the analysed posts showed positive sentiments, mostly associated with terms related to sustainability or climate action (> 13%).

planning of awareness campaigns, announcements, or events⁵⁵. Such is particularly relevant when social media data is gathered in near real-time, allowing governments and non-governmental organizations to monitor shifts in public sentiments and quickly develop communication strategies that resonate with the prevailing emotions, providing timely support and fostering resilience-building efforts⁵⁶. Furthermore, exploring the most salient extreme weather events may guide disaster preparedness planning and infrastructure development to mitigate the effects of these phenomena⁵⁷.

Even though our study has only been conducted in a specific test area, we are confident that it serves as a case study that helps advance the ability of using pre-trained and open-source deep learning models to support the social monitoring of climate change from textual content shared on social media platforms. Our methodological approaches (sentiment and emotion analysis, irony detection, and temporal pattern analysis of social media content) are based on pre-trained and open-source deep learning models that are inherently adaptable. With the continuous improvement of multilingual NLP tools, including BERT models, the framework can be fine-tuned to account for linguistic and cultural differences, allowing its application to a wide variety of regions. Climate change is a global issue, and many of its impacts, such as extreme weather events, resonate universally. The emotional and perceptual patterns identified in our study, such as the prevalence of negativity associated with disasters, may similarly emerge in other regions, providing valuable insights into public perceptions of climate change⁵⁸. However, the generalization of our approach requires careful consideration of cultural and linguistic nuances, as emotions and sentiments are deeply influenced by cultural factors, and their expression may vary significantly between regions. Adapting our approach would require fine-tuning models to account for such differences, ensuring accurate sentiment, emotion and irony detection.

Methods

Test area

The test area is Southwestern Europe which includes the territories of Portugal and Spain (mainland and islands). This test area was selected to serve as a microcosm for studying broader climate change trends in Europe and beyond, providing valuable insights into regional variations, and facilitating cross-border collaboration on climate-related initiatives. Southwestern Europe climate is highly vulnerable to variations in precipitation and temperature patterns and are particularly vulnerable to the impacts of extreme weather events⁵⁹. For instance, rising sea levels threaten coastal communities and infrastructure, while changing precipitation patterns affect agriculture and water resources⁶⁰.

Methodological framework

We deployed a five-step methodological framework. Firstly, we collected posts on climate change and related phenomena for three languages—Portuguese, Spanish and English—using the Application Programming Interface (API) from X (Twitter) under an academic research licence. After removing duplicates, we trained and evaluated one natural language processing (NLP) model for each language, to exclude irrelevant posts. The final dataset was then analysed to identify temporal trends and patterns in the scope of climate change. Finally,

we performed two perceptions analyses - sentiment and emotion, and one expressions detection - irony, using openly available pre-trained deep learning models. All steps were implemented using Google Colaboratory (Colab), a free Jupyter notebook environment from Google Research that allows the execution of Python code entirely in the cloud, requiring no setup to use and providing free access to computing resources, including GPUs (<https://colab.research.google.com/notebooks/intro.ipynb>). The training, evaluation and remaining analyses were performed using free and open-source platforms specialised in the training and inference of natural language processing models⁶¹. Specifically, we used Keras (<https://keras.io/>), TensorFlow (<https://www.tensorflow.org/>), spaCy (<https://spacy.io/>) and Scikit-learn (<https://scikit-learn.org/stable/>) libraries, along with Transformers from Hugging Face (<https://huggingface.co/>) and pysentimiento (<https://github.com/pysentimiento/pysentimiento/>).

Data collection

We compiled a dataset of X textual content on climate change and related phenomena, impacts and effects. First, we defined a set of more than 130 potentially relevant keywords (e.g., wildfires, global warming, floods, pollution), in Portuguese, Spanish and English languages (see Tables S1, S2 and S3), with the support of native speakers, who were also experts in the field of climate research. We collected the textual data (except retweets) using Twitter’s API v2 academic research licence along with a publicly available open-source TwitterAPI Python library: Tweepy (<https://www.tweepy.org/>), from March 2006 (i.e., since the creation of Twitter) to the end of 2022 (which resulted in 174,674 posts for Portuguese, 1,224,504 posts for Spanish and 372,655 posts for English language). Using the TwitterAPI, we restricted the search query to the territories of Portugal and Spain. Searches were performed considering both singular and plural forms of each keyword, as well as an initial string of characters and a wildcard (i.e., character or sequence of characters that acts as a placeholder, allowing for flexible matching in searches or patterns). Metadata, such as post’s unique identifier, date of creation, location, bounding box, username, and user location, were also gathered, alongside the textual content. Data collection procedure was optimized to extract only the essential data, from openly available posts. This allowed mitigating the environmental footprint of digital analysis, while respecting the user’s privacy⁶² and the platform’s terms of use in effect at the date of collection.

Data preprocessing

We removed duplicate posts, as well as posts with less than six words, given that longer posts may contain more detailed and informative content, offering better context to understand sentiments and enhancing language model training (see Supplementary Material for details). Additionally, we implemented the “preprocessing” function of the pysentimiento library (<https://github.com/pysentimiento/pysentimiento/>) to clean the textual content of posts. This function included replacing user handles and URLs by special tokens, shortening of repeated characters, normalizing laughter expressions, and processing hashtags and emojis.

Data annotation

We manually annotated a random subsample of around 1200 posts (42% relevant and 58% irrelevant) for Portuguese and English and 3200 (47% relevant and 53% irrelevant) for Spanish, to exclude those not related to climate change impacts and phenomena. The posts’ text was classified as “Relevant” if the content was directly related to climate phenomena and impacts, or as “Irrelevant” if the opposite was verified – if the content was unclear and/or not related to climate change (Table 3). Since the two classes in each random subset (Portuguese, Spanish and English) were not balanced between “Relevant” and “Irrelevant” classes, we implemented a balancing technique through empirical trials—undersampling—to prevent bias in the model performance. This technique consisted in keeping all posts in the minority class (“Relevant”) while decreasing the size of the majority class (“Irrelevant”).

Model selection and parametrization

For each language, we used open-source BERT (Bidirectional Encoder Representations from Transformers) models: BERTimbau for Portuguese⁶³, BETO for Spanish⁶⁴, and BERT base⁶⁵ for English. Advantages of these models are their ease of transfer learning and high performance in similar tasks^{66,67}. For model optimization, we used the Adam optimizer algorithm⁶⁸ and set the batch size to 16, for BERTimbau, and 4, for BETO and BERT base. Adam optimizer efficiently navigates the model’s high-dimensional parameter space, with the aim of finding a set of weights that minimises the loss function, and thus increasing the model’s performance⁶⁸. The adaptive learning rates and momentum features of Adam contribute to the optimization process’s effectiveness

	Relevant	Irrelevant
Portuguese	“Seca: o Algarve, no sul de Portugal, só tem água para um ano. Se não chover, poderá chegar ao fim de 2023 sem água para o público #seca”	“Que seca, tenho sair de casa mas tá tudo a trabalhar”
Spanish	“Estamos viviendo momentos extraños. Acaba de pasar por #Granada una #gotafria #TormentaElectrica, como jamás había visto y sentido”	“Fuieste la calma que precedió a la tormenta, como el que se santigua sabiendo que va a la guerra”
English	“The smell outside right now of wildfires — it’s so intense and this is kms away”	“I don’t care if y’all get triggered, editing film photos with the infrared look is kinda fire”

Table 3. Examples of posts considered “relevant” and “irrelevant” by the models in Portuguese, Spanish and english.

Metric	Definition
Accuracy	The proportion of correctly classified posts as “Relevant” and as “Irrelevant” by the model
Precision	The proportion of correctly classified posts as “Relevant” by the model, in relation to all actual and predicted “Relevant” posts
Recall	The proportion of correctly classified posts as “Relevant” by the model, in relation to all actual “Relevant” posts
F1-score	The harmonic mean of Precision and Recall with respect to the “Relevant” class

Table 4. Metrics used for the evaluation of the natural processing models, with corresponding examples.

and stability. Both epochs and learning rates were chosen through empirical trials, with 20 and $2e^{-6}$, respectively, showing the best performance for the three models.

Model training, validation, and evaluation

The dataset for each language was split into training and testing sets, with 75% allocated for training and 25% for testing. A fixed seed of seven was used to ensure reproducibility across runs. This was achieved by preserving the class distribution in both the training and testing sets through stratified sampling, ensuring a similar proportion of each class in both sets. This approach prevents any imbalance in the categories, which is crucial for tasks with skewed class distributions. During training, each model (and language) was then validated by retaining 10% of the training posts for validation (Tables S4, S5 and S6). Additionally, to further validate the model’s performance and avoid overfitting, we saved the best model based on validation loss, ensuring the use of the optimal model throughout the training process. The performance of each model (BERTimbau, BETO and BERT base) was assessed based on common classification metrics (Table 4⁶⁹: accuracy (ACC), precision (positive predictive value, PPV), recall (sensitivity or true positive rate: TPR) and F1-score (F_1). The term positive stands for posts classified as “Relevant”, whereas negative stands for the opposite “Irrelevant”. Lastly, we removed the posts predicted by the models as “Irrelevant”, resulting in a final set of 38,419 posts for Portuguese, 569,610 posts for Spanish and 61,701 posts for English, which was considered in following analyses.

Time series and frequency analyses

We implemented a time series analysis, which consisted of calculating the number of posts over time (2010 to 2022), allowing the identification of potential frequency peaks. To ensure the significance of the detected peaks, we performed an anomaly detection analysis, on which we implemented the Isolation Forest algorithm from the Scikit-learn library (<https://scikit-learn.org/stable/>).

We also used spaCy modules (Attributes, <https://spacy.io/api/attributes>) and classes (Matcher, <https://spacy.io/api/matcher>) for the three languages to develop frequency functions based on regex (regular expressions) and lemma (lemmatization) matching attributes. The regex function enables precise text pre-processing by filtering out non-alphabetic characters and symbols, ensuring a cleaner input for analysis, while the lemma function plays a crucial role in standardising word forms, reducing them to their base or dictionary forms⁷⁰. The synergy between regex and lemma functions not only reduces noise but also streamlines the tokenization process, providing a more coherent and normalised basis for computing word frequencies⁷⁰. Then, we implemented the frequency functions for each phenomenon across time. To explore the three most common phenomena in the detected peaks, we applied the same frequency functions for each peak (i.e., based on spaCy modules and classes). To avoid repetition and saturated charts, we developed the visualisations considering the classes in the “Specific group” level (Tables S1, S2 and S3). For the two analyses (time series and frequency), we aggregated the results from the three languages, in order to facilitate the interpretation and understanding of the outputs across the test area.

Perceptions and expressions analysis

We tested three different perception and expression tasks (one for each language): sentiment, emotion and irony. Although some previously established frameworks consider emotions as subsets of sentiments, we decided to separate sentiment and emotion analysis to provide a more nuanced understanding of social media discourse on climate change. While sentiments provide a broad perspective, emotions offer finer granularity, allowing to capture the underlying emotional states associated with sentiments. Ultimately, this approach enables to highlight trends that might be masked when emotions are aggregated into broader sentiment categories. The three tasks were performed using the same approach, which consisted of using nine pre-trained transformer-based models, available through the pysentimiento library (<https://github.com/pysentimiento/pysentimiento/>). These transformer-based models, which use contextual embeddings, were pre-trained on datasets similar to ours, to learn evaluative language patterns (sentiment analysis), emotional expressions markers (emotion analysis) and linguistic inconsistencies or contrasts (irony analysis), for each language. For ease of interpretation, we aggregated the results from the three languages. To evaluate the performance of the models, a random sample of 3000 posts (1000 per language) was manually annotated by two researchers (with inter-annotator agreement values above 94%, Tables S7 and S8) following the same classification established by the models (“positive”, “neutral” and “negative” for the sentiment analysis; “anger”, “joy”, “disgust”, “surprise”, “fear” and “neutral” for the emotion analysis; “ironic” and “not ironic” for the irony detection).

Sentiment analysis

Sentiment analysis, also known as opinion mining, is an NLP technique used to determine the emotional tone and attitude expressed in textual content⁴⁸. This enables the quantification of subjective information, providing significant insights into public opinions and attitudes. In this task, we implemented a multiclass classifier for the three languages that returned a single variable (positive - “POS”, neutral - “NEU” or negative - “NEG”), depending on whether the sentiment/tone implicit in the post was positive, neutral, or negative. This analysis was performed for all tweets across time, as well as for each specific phenomenon, to fully investigate the trends in users’ perceptions (see Table S9 for examples of predicted posts).

Emotion analysis

Emotion analysis is also an NLP task that focuses on the identification, extraction, and interpretation of human emotions and attitudes in textual content, offering valuable knowledge of people’s feelings and emotions⁷¹. This task was performed using a multiclass classifier for the English and Spanish languages, that also returned a single variable (“anger”, “joy”, “sadness”, “fear”, “surprise”, “disgust” or “neutral”), and a multilabel classifier for the Portuguese language, that returned a list of labels (e.g., “admiration”, “confusion”, “disapproval”, “fear”, “joy”, and “love”, among others). To ensure cross-language comparability, we decided to only consider the six emotions proposed in Ekman’s basic emotion model (“anger”, “joy”, “sadness”, “fear”, “surprise” and “disgust”)⁷², along with the “neutral” category. The “neutral” category functions as a default or non-emotional classification, encompassing posts that either lack any discernible emotion or contain an emotion that does not align with the six basic emotions. The inclusion of this category prevents the model from forcing posts to fit into one of the other emotion categories, ensuring a more accurate and meaningful classification (see Table S10 for examples of predicted posts). While this emotion analysis approach may reduce granularity, it was necessary for standardizing categories across the three languages. Ekman’s model is widely used in emotion research and provides a well-established framework for comparative analysis⁷². Previous studies have shown that fine-grained emotion labels can be effectively mapped to these six basic categories based on their affective dimensions^{73,74}. This standardization not only ensures comparability across languages but also enhances the interpretability of the results by focusing on universally recognized emotional expressions.

Irony detection

Irony detection refers to the process of identifying and distinguishing instances of irony, sarcasm, or other forms of figurative language in textual content⁷⁵. This task was also relevant to include in our study, as accurate irony detection can enhance sentiment and emotion analysis by ensuring that the true sentiment of a statement is captured, avoiding misinterpretations of the text. Unlike the previous tasks, for irony detection we implemented a binary classifier that generated a single binary variable (“ironic” or “not ironic”; see Table S11 for examples of predicted posts). One of the main challenges in irony detection is accounting for culturally specific expressions of irony, as these may vary significantly across languages and social groups⁷⁶. To overcome this, we decided to use a high-performance binary classifier that was pre-trained and fine-tuned on datasets annotated for irony detection in each specific language, ensuring exposure to a broad range of ironic expressions. Despite this, we acknowledge that the inherent subjectivity and variability of irony may still pose limitations, particularly for highly implicit or context-dependent cases.

Data availability

The dataset compiled in the current study is not publicly available due to privacy and ethical concerns, such as potential exposure of personal information and misinterpretation of context but may be partially available (after anonymization) from the corresponding author on reasonable request. All code and models developed and implemented in this research are available in GitHub: <https://github.com/anascardoso/Climatemediaproject> and the following website: <https://climatemediaproject.wordpress.com/>.

Received: 5 July 2024; Accepted: 4 April 2025

Published online: 28 April 2025

References

1. Austin, E. et al. Concerns about climate change among rural residents in Australia. *J. Rural Stud.* **75**, 98–109. <https://doi.org/10.1016/j.jrurstud.2020.01.010> (2020).
2. Nations, U. Global sustainable development report 2023—times of crisis, times of change: science for accelerating transformations to sustainable development (2023).
3. Leonard, M., Pisani-Ferry, J., Shapiro, J., Tagliapietra, S. & Wolff, G. B. *The Geopolitics of the European Green Deal* (Bruegel policy contribution, 2021). <https://www.econstor.eu/bitstream/10419/237660/1/1749375737.pdf>.
4. Benansio, J. S. et al. Perceptions and attitudes towards climate change in fishing communities of the Sudd wetlands, South Sudan. *Reg. Environ. Chang.* **22**(2), 78. <https://doi.org/10.1007/s10113-022-01928-w> (2022).
5. Bechtoldt, M. N., Götmann, A., Mosler, U. & Pauw, W. P. Addressing the climate change adaptation puzzle: a psychological science perspective. *Clim. Policy.* **21**(2), 186–202. <https://doi.org/10.1080/14693062.2020.1807897> (2021).
6. Hornsey, M. J., Harris, E. A., Bain, P. G. & Fielding, K. Meta-analyses of the determinants and outcomes of belief in climate change. *Nat. Clim. Change.* **6**(6), 622–626. <https://doi.org/10.1038/nclimate2943> (2016).
7. Van der Linden, S. The social-psychological determinants of climate change risk perceptions: towards a comprehensive model. *J. Environ. Psychol.* **41**, 112–124. <https://doi.org/10.1016/j.jenvp.2014.11.012> (2015).
8. Sisco, M. R., Bosetti, V. & Weber, E. U. When do extreme weather events generate attention to climate change? *Clim. Change.* **143**, 227–241. <https://doi.org/10.1007/s10584-017-1984-2> (2017).
9. Bradley, G. L., Babutsidze, Z., Chai, A. & Reser, J. P. The role of climate change risk perception, response efficacy, and psychological adaptation in pro-environmental behavior: a two Nation study. *J. Environ. Psychol.* **68**, 101410. <https://doi.org/10.1016/j.jenvp.2020.101410> (2020).

10. Willaert, T., Van Eecke, P., Beuls, K. & Steels, L. Building social media observatories for monitoring online opinion dynamics. *Social Media + Soc.* **6**(2), 2056305119898778. <https://doi.org/10.1177/2056305119898778> (2020).
11. Pedrero-Esteban, L. M. & Barrios-Rubio, A. Digital communication in the age of immediacy. *Digital* **4**(2), 302–315. <https://doi.org/10.3390/digital4020015> (2024).
12. Brosch, T. Affect and emotions as drivers of climate change perception and action: a review. *Curr. Opin. Behav. Sci.* **42**, 15–21. <https://doi.org/10.1016/j.cobeha.2021.02.001> (2021).
13. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proc. Natl. Acad. Sci.* **118**(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118> (2021).
14. Correia, R. et al. Inferring public interest from search engine data requires caution. *Front. Ecol. Environ.* **17**, 5. <https://doi.org/10.1002/fee.2048> (2019).
15. Chen, K. et al. How climate movement actors and news media frame climate change and strike: evidence from analyzing twitter and news media discourse from 2018 to 2021. *Int. J. Press/Politics* **28**(2), 384–413. <https://doi.org/10.1177/19401612221106405> (2023).
16. Su, L. Y. F., Akin, H. & Brossard, D. Methods for assessing online climate change communication, social media discussion, and behavior. In *Oxford Research Encyclopedia of Climate Science*. <https://doi.org/10.1093/acrefore/9780190228620.013.492> (2017).
17. Bishop, L. & Gray, D. Ethical challenges of publishing and sharing social media research data. In *The Ethics of Online Research*, vol. 2 159–187 (Emerald Publishing Limited, 2017). <https://doi.org/10.1108/S2398-60182018000002007>.
18. Upadhyaya, A., Fisichella, M. & Nejd, W. A. Multi-task model for sentiment aided stance setection of climate change tweets. In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 17 854–865 (2023). <https://doi.org/10.1609/icws.m.v17i1.22194>.
19. Dahal, B., Kumar, S. A. & Li, Z. Topic modeling and sentiment analysis of global climate change tweets. *Social Netw. Anal. Min.* **9**, 1–20. <https://doi.org/10.1007/s13278-019-0568-8> (2019).
20. Bryan-Smith, L. et al. Real-time social media sentiment analysis for rapid impact assessment of floods. *Comput. Geosci.* **178**, 105405. <https://doi.org/10.1016/j.cageo.2023.105405> (2023).
21. Karimiziarani, M. & Moradkhani, H. Social response and disaster management: insights from Twitter data assimilation on hurricane Ian. *Int. J. Disaster Risk Reduct.* **95**, 103865. <https://doi.org/10.1016/j.ijdrr.2023.103865> (2023).
22. Lydiri, M., El Mourabit, Y., Habouz, E., Fakir, M. & Y. A. A performant deep learning model for sentiment analysis of climate change. *Social Netw. Anal. Min.* **13**(1), 8. <https://doi.org/10.1007/s13278-022-01014-3> (2022).
23. Loureiro, M. L. & Alló, M. Sensing climate change and energy issues: sentiment and emotion analysis with social media in the UK and Spain. *Energy Policy* **143**, 111490. <https://doi.org/10.1016/j.enpol.2020.111490> (2020).
24. Fagbola, T. M., Abayomi, A., Mutanga, M. B. & Jugoo, V. Lexicon-based sentiment analysis and emotion classification of climate change related tweets. In *International Conference on Soft Computing and Pattern Recognition* 637–646 (Springer International Publishing, 2022). https://doi.org/10.1007/978-3-030-96302-6_60.
25. Andrade, C., Contente, J. & Santos, J. A. Climate change projections of dry and wet events in Iberia based on the WASP-Index. *Climate* **9**(6), 94. <https://doi.org/10.3390/cli9060094> (2021).
26. Pino, D. et al. Meteorological and hydrological analysis of major floods in NE Iberian Peninsula. *J. Hydrol.* **541**, 63–89. <https://doi.org/10.1016/j.jhydrol.2016.02.008> (2016).
27. Moemken, J. & Pinto, J. G. Recurrence of drought events over Iberia. Part i: methodology and application for present climate conditions. *Tellus A: Dyn. Meteorol. Oceanogr.* **74**, 456. <https://doi.org/10.16993/tellusa.50> (2022).
28. Nash, N., Capstick, S., Whitmarsh, L., Chaudhary, I. & Manandhar, R. Perceptions of local environmental issues and the relevance of climate change in Nepal's Terai: perspectives from two communities. *Front. Sociol.* **4**, 60. <https://doi.org/10.3389/fsoc.2019.00060> (2019).
29. Owen, A. L., Conover, E., Videras, J. & Wu, S. Heat waves, droughts, and preferences for environmental policy. *J. Policy Anal. Manag.* **31**(3), 556–577. <https://doi.org/10.1002/pam.21599> (2012).
30. Vaz, A. S. et al. Perceptions of public officers towards the effects of climate change on ecosystem services: A Case-Study from Northern Portugal. *Front. Ecol. Evol.* **9**, 710293. <https://doi.org/10.3389/fevo.2021.710293> (2021).
31. Pinto, P. et al. Influence of convectively driven flows in the course of a large fire in Portugal: the case of Pedrógão Grande. *Atmosphere* **13**(3), 414. <https://doi.org/10.3390/atmos13030414> (2022).
32. Ramos, A. M. et al. The compound event that triggered the destructive fires of October 2017 in Portugal. *Iscience* **26**, 3. <https://doi.org/10.1016/j.isci.2023.106141> (2023).
33. Rodrigues, M. et al. Drivers and implications of the extreme 2022 wildfire season in Southwest Europe. *Sci. Total Environ.* **859**, 160320. <https://doi.org/10.1016/j.scitotenv.2022.160320> (2023).
34. Gonçalves, A. C., Nieto, R. & Liberato, M. L. Synoptic and dynamical characteristics of high-impact storms affecting the Iberian Peninsula during the 2018–2021 extended winters. *Atmosphere* **14**(9), 1353. <https://doi.org/10.3390/atmos14091353> (2023).
35. Diaz-Hernandez, G., Mendez, F. J. & Minguez, R. Numerical analysis and diagnosis of the hydrodynamic effects produced by hurricane Gordon along the Coast of Spain. *Weather Forecast.* **29**(3), 666–683. <https://doi.org/10.1175/WAF-D-13-00130.1> (2014).
36. Gu, M., Guo, H., Zhuang, J., Du, Y. & Qian, L. Social media user behavior and emotions during crisis events. *Int. J. Environ. Res. Public Health.* **19**(9), 5197. <https://doi.org/10.3390/ijerph19095197> (2022).
37. Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S. & Danforth, C. M. Climate change sentiment on Twitter: an unsolicited public opinion poll. *PloS One.* **10**(8), e0136092. <https://doi.org/10.1371/journal.pone.0136092> (2015).
38. Bergquist, M., Nilsson, A. & Schultz, P. Experiencing a severe weather event increases concern about climate change. *Front. Psychol.* **10**, 220. <https://doi.org/10.3389/fpsyg.2019.00220> (2019).
39. Dong, Z. S., Meng, L., Christenson, L. & Fulton, L. Social media information sharing for natural disaster response. *Nat. Hazards* **107**, 2077–2104. <https://doi.org/10.1007/s11069-021-04528-9> (2021).
40. Tyson, A., Kennedy, B., Funk, C. & Gen, Z. Millennials stand out for climate change activism, social media engagement with issue. *Pew Res. Cent.* **2021**, 26 (2021).
41. Becker, A. & Anderson, A. A. Using humor to engage the public on climate change: the effect of exposure to one-sided vs. two-sided satire on message discounting, elaboration and counterarguing. *J. Sci. Commun.* **18**(4), A07. <https://doi.org/10.22323/2.18040207> (2019).
42. Paz, M. A., Mayagoitia-Soria, A. & González-Aguilar, J. M. From polarization to Hate: portrait of the Spanish political meme. *Social media + Soc.* **7**(4), 20563051211062920. <https://doi.org/10.1177/20563051211062920> (2021).
43. Akhtyrskaya, K. Linguistic expression of irony in social media (2014). <http://hdl.handle.net/10400.1/8365>.
44. Boot, A. B., Kim Sang, T., Dijkstra, E., Zwaan, R. A. & K., & How character limit affects Language usage in tweets. *Palgrave Commun.* **5**, 1. <https://doi.org/10.1057/s41599-019-0280-3> (2019).
45. Rosenberg, E. et al. Sentiment analysis on Twitter data towards climate action (2023). <https://doi.org/10.1016/j.rineng.2023.101287>.
46. Taylor, N. G. et al. The future for mediterranean wetlands: 50 key issues and 50 important conservation research questions. *Reg. Environ. Chang.* **21**, 1–17. <https://doi.org/10.1007/s10113-020-01743-1> (2021).
47. Hargittai, E. Potential biases in big data: omitted voices on social media. *Social Sci. Comput. Rev.* **38**(1), 10–24. <https://doi.org/10.1177/0894439318788322> (2020).
48. Liu, B. et al. Context-aware social media user sentiment analysis. *Tsinghua Sci. Technol.* **25**(4), 528–541. <https://doi.org/10.26599/TST.2019.9010021> (2020).

49. Wankhade, M., Rao, A. C. S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* **55**(7), 5731–5780. <https://doi.org/10.1007/s10462-022-10144-1> (2022).
50. Wilkins, E. J., Howe, P. D. & Smith, J. W. Social media reveal ecoregional variation in how weather influences visitor behavior in US National park service units. *Sci. Rep.* **11**(1), 2403. <https://doi.org/10.1038/s41598-021-82145-z> (2021).
51. Jane, S. F. et al. News media and fisheries-independent data reveal hidden impacts of hurricanes. *Ambio* **51**(10), 2169–2181. <https://doi.org/10.1007/s13280-022-01732-0> (2022).
52. Zhang, B. et al. Changes in public sentiment under the background of major emergencies—taking the Shanghai epidemic as an example. *Int. J. Environ. Res. Public Health*. **19**(19), 12594. <https://doi.org/10.3390/ijerph191912594> (2022).
53. Funk, C. & Hefferon, M. US public views on climate and energy. *Pew Res. Cent.* **2019**, 25 (2019).
54. Eslen-Ziya, H. Humour and sarcasm: expressions of global warming on Twitter. *Humanit. Social Sci. Commun.* **9**(1), 1–8. <https://doi.org/10.1057/s41599-022-01236-y> (2022).
55. Wei, Y., Gong, P., Zhang, J. & Wang, L. Exploring public opinions on climate change policy in big data Era—a case study of the European union emission trading system (EU-ETS) based on Twitter. *Energy Policy*. **158**, 112559. <https://doi.org/10.1016/j.enpol.2021.112559> (2021).
56. Veltri, G. A. & Atanasova, D. Climate change on Twitter: content, media ecology and information sharing behaviour. *Public. Underst. Sci.* **26**(6), 721–737. <https://doi.org/10.1177/0963662515613702> (2017).
57. Raymond, C. et al. Understanding and managing connected extreme events. *Nat. Clim. Change* **10**(7), 611–621. <https://doi.org/10.1038/s41558-020-0790-4> (2020).
58. Capstick, S., Whitmarsh, L., Poortinga, W., Pidgeon, N. & Upham, P. International trends in public perceptions of climate change over the past quarter century. *Wiley Interdiscipl. Rev.: Clim. Change*. **6**(1), 35–61. <https://doi.org/10.1002/wcc.321> (2015).
59. Fernández-Palacios, J. M. et al. Climate change and human impact in Macaronesia (2016). <https://doi.org/10.22498/pages.24.2.68>.
60. Torres, C. et al. Climate change and its impacts in the Balearic Islands: a guide for policy design in mediterranean regions. *Reg. Environ. Chang.* **21**(4), 107. <https://doi.org/10.1007/s10113-021-01810-1> (2021).
61. Abadi, M. et al. {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* 265–283 (2016). <https://doi.org/10.48550/arXiv.1605.08695>.
62. Di Minin, E., Fink, C., Hausmann, A., Kremer, J. & Kulkarni, R. How to address data privacy concerns when using social media data in conservation science. *Conserv. Biol.* **35**(2), 437–446. <https://doi.org/10.1111/cobi.13708> (2021).
63. Souza, F., Nogueira, R. & Lotufo, R. BERTimbau: pretrained BERT Models for Brazilian Portuguese. In *Intelligent Systems. BRACIS 2020. Lecture Notes in Computer Science*, vol 12319 (eds. Cerri, R. & Prati, R.C.) https://doi.org/10.1007/978-3-030-61377-8_28 (2020).
64. Wu, S. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. arXiv preprint arXiv:1904.09077. <https://doi.org/10.48550/arXiv.1904.09077> (2019).
65. Devlin, J., Chang, M. W., Lee, K., Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 <https://doi.org/10.48550/arXiv.1810.04805> (2018).
66. Varini, F. S., Boyd-Graber, J., Ciaramita, M. & Leippold, M. ClimaText: a dataset for climate change topic detection. arXiv preprint arXiv:2012.00483. <https://doi.org/10.48550/arXiv.2012.00483> (2020).
67. Effrosynidis, D., Karasakalidis, A. I., Sylaios, G. & Arampatzis, A. The climate change Twitter dataset. *Expert Syst. Appl.* **204**, 117541. <https://doi.org/10.1016/j.eswa.2022.117541> (2022).
68. Kingma, D. P., Ba, J. L. & Adam A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings* 1–15 (2015).
69. Tharwat, A. Classification assessment methods. *Appl. Comput. Inf.* **17**(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003> (2018).
70. Basha, S. M. & Fathima, A. S. *Natural Language Processing: Practical Approach* (MileStone Research, 2023).
71. Rout, J. K. et al. A model for sentiment and emotion analysis of unstructured social media text. *Electron. Commer. Res.* **18**, 181–199. <https://doi.org/10.1007/s10660-017-9257-8> (2018).
72. Ekman, P. Are there basic emotions? *Psychol. Rev.* **99**(3), 550–553. <https://doi.org/10.1037/0033-295X.99.3.550> (1992).
73. Demszky, D. et al. GoEmotions: A dataset of fine-grained emotions. arXiv preprint arXiv:2005.00547. <https://doi.org/10.48550/arXiv.2005.00547> (2020).
74. Sam Abraham, S., Gangan, P. & VL, L., & M Readers' affect: predicting and Understanding readers' emotions with deep learning. *J. Big Data* **9**(1), 1–31. <https://doi.org/10.1186/s40537-022-00614-2> (2022).
75. Ghosh, A. & Veale, T. Fracking sarcasm using neural network. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* 161–169 (2016). <https://doi.org/10.18653/v1/W16-0425>.
76. Ghanem, B., Karoui, J., Benamara, F., Rosso, P. & Moriceau, V. Irony detection in a multilingual context. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II* 42 141–149 (Springer International Publishing, 2020). https://doi.org/10.1007/978-3-030-45442-5_18.

Acknowledgements

This study was supported by FCT (Portuguese Foundation for Science and Technology) through project ClimateMedia – Understanding climate change phenomena and impacts from digital technology and social media [contract reference 2022.06965.PTDC; doi: 10.54499/2022.06965.PTDC]. ASC was supported by the 2021 PhD Research Studentships [grant reference 2021.05426.BD; doi: 10.54499/2021.05426.BD]; ASR was supported by a Ramón y Cajal fellowship [contract reference RYC2023-043755-I] from MICIU/AEI/10.13039/501100011033 and FSE+; JS was supported by FCT, under the project UIDB/04033/2020; ASV was supported by FCT's program for Stimulus for Scientific Employment – Individual Support [contract reference 2020.01175.CEECIND/CP1601/CT0009; doi: 10.54499/2020.01175.CEECIND/CP1601/CT0009]. The authors thank Dr. Nuria Pistón Caballero, from the Universidad de Granada (Spain) for advice and recommendations on the selection of Spanish keywords.

Author contributions

ASC was responsible for data collection, analysis and writing of the main manuscript text. CdS helped with data collection and literature review. ASR, IJ and SB contributed to data interpretation. JAS also helped with data collection and data interpretation. AJ led the methodological supervision, review; ASV was responsible for the ideation and the research supervision. All authors reviewed the manuscript.

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/s41598-025-97441-1>.

Correspondence and requests for materials should be addressed to A.S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025