




Artificial intelligence for modeling and understanding extreme weather and climate events

Received: 13 June 2024

Accepted: 23 January 2025

Published online: 24 February 2025

 Check for updates

Gustau Camps-Valls ¹✉, Miguel-Ángel Fernández-Torres ¹, Kai-Hendrik Cohrs ¹, Adrian Höhl ², Andrea Castelletti ^{3,4}, Aytac Pacal ^{5,6}, Claire Robin ^{7,8}, Francesco Martinuzzi ^{9,10}, Ioannis Papoutsis ^{11,12,13}, Ioannis Prapas ^{1,11,12}, Jorge Pérez-Aracil ¹⁴, Katja Weigel ^{5,6}, Maria Gonzalez-Calabuig ¹, Markus Reichstein ^{7,8}, Martin Rabel ¹⁵, Matteo Giuliani ³, Miguel D. Mahecha ^{9,10}, Oana-Iuliana Popescu ¹⁵, Oscar J. Pellicer-Valero ¹, Said Ouala ¹⁶, Sancho Salcedo-Sanz ¹⁴, Sebastian Sippel ¹⁷, Spyros Kondylatos ^{1,11,12}, Tamara Happé ¹⁸ & Tristan Williams ¹

In recent years, artificial intelligence (AI) has deeply impacted various fields, including Earth system sciences, by improving weather forecasting, model emulation, parameter estimation, and the prediction of extreme events. The latter comes with specific challenges, such as developing accurate predictors from noisy, heterogeneous, small sample sizes and data with limited annotations. This paper reviews how AI is being used to analyze extreme climate events (like floods, droughts, wildfires, and heatwaves), highlighting the importance of creating accurate, transparent, and reliable AI models. We discuss the hurdles of dealing with limited data, integrating real-time information, and deploying understandable models, all crucial steps for gaining stakeholder trust and meeting regulatory needs. We provide an overview of how AI can help identify and explain extreme events more effectively, improving disaster response and communication. We emphasize the need for collaboration across different fields to create AI solutions that are practical, understandable, and trustworthy to enhance disaster readiness and risk reduction.

The frequency, intensity, and duration of climate extremes have increased in recent years, posing unprecedented challenges to societal stability, economic security, biodiversity loss, and ecological integrity¹. These events—ranging from severe storms and floods to droughts and heatwaves—exert profound impacts on human livelihoods and the natural environment, often with long-lasting and sometimes irreversible consequences. Modeling, characterizing, and understanding extreme events is key to advancing mitigation and adaptation strategies.

Providing a precise, formal, universally applicable definition of an extreme event is inherently difficult. Statistical definitions that define

extremes as the tails of a probability distribution (e.g., the 1% or 5%) are often insufficient as they may fail to account for complex multi-dimensional interdependencies¹, such as changes in societal and geographic context and the cumulative nature of impacts. In this context, Artificial Intelligence (AI) has emerged as a transformative tool for the detection², forecasting^{3,4}, analysis of extreme events, and generation of worst-case events⁵, and promises advances in attribution studies, explanation, and communication of risk⁶. The capabilities of machine learning (ML) and deep learning (DL), in particular, in combination with computer vision techniques, are advancing the detection and

A full list of affiliations appears at the end of the paper. ✉ e-mail: gustau.camps@uv.es

localization of events by exploiting climate data, such as reanalysis and observations⁷. Modern techniques for quantifying uncertainty⁸ are necessary to progress in climate change risk assessment⁹. Using ensembles in combination with AI models has progressed the field of attribution of extremes¹⁰, identification of patterns, trends¹¹ and climate analogs¹². Yet, AI does not only excel at prediction but can also explain processes (e.g., via explainable AI (XAI)¹³ and causal inference¹⁴), which is essential for decision-making and effective mitigation strategies. Recent advances in large language models (LLMs) that retrieve information from heterogeneous text sources allow for effective integration of methods for communication tasks and human-machine interaction in situation analysis¹⁵.

Recent reviews have explored AI applications for extreme weather and climate events, covering topics like deep learning for medium-range forecasts, sub-seasonal to decadal predictions, and causality and explainability in extreme atmospheric events^{7,16,17}. However, they lack a holistic view that includes the broader environmental and societal impacts of these events. Here, we review the role of AI in extreme event analysis, its challenges and opportunities. The general pipeline of AI-driven extreme event analysis (cf. Fig. 1) encapsulates the entire workflow from data collection and preprocessing to the generation of outputs such as predictions, patterns, trends, climate attributions, and causal relations. The framework also highlights the iterative nature of AI processes, where outputs serve their direct purpose as well as to inform and improve data collection, preprocessing, and hypothesis formulation.

We review AI for modeling, detection, forecasting, and communication (“Review of AI methods”), and discuss key challenges related to data and models but also to effective risk communication and system’s integration (“Data, model and integration challenges”). We present case studies on AI applications for droughts, heatwaves, wildfires, and floods (“Case Studies”) and conclude by outlining future research opportunities in AI for extreme events (“Conclusions and Outlook”).

Review of AI methods

This section reviews the main methods in all aspects of extreme event analysis (cf. Table 1): data and models, understanding and trustworthiness, and the last mile on communication and deployment. A glossary of terms is provided in Table S1, while a list of strengths and weaknesses of different ML models is given in Table S2 in the Supplementary material.

Extreme event modeling

AI methodologies for extreme event modeling can be categorized into detection¹⁸, prediction¹⁹, and impact assessment²⁰ tasks. Thanks to the

emergence and success of DL, all of these tasks can be tackled by designing data-driven models that exploit spatio-temporal and multi-source Earth data characteristics, from climate variables to in situ measurements and satellite remote sensing images (see Fig. 2 [top row]).

Detection. Detecting and identifying extreme events geographically over time is fundamental for assessing impacts and improving anticipation and mitigation strategies. Also, detection is the first step in discovering underlying patterns and better understanding their generating processes and mechanisms.

Classical statistical methods, such as threshold or percentile-based indices, have been widely applied to detect extreme events (see Table 1). These methods often miss events identified by experts or impact-based detection approaches²¹ and fail to capture event complexity, as they focus on single variables and overlook hazard thresholds that vary across space and time²². To address this limitation, AI methodologies can help reconcile expert knowledge with data-driven approaches as they capture regularities and nuanced relations in large volumes of observational data, consider multiple variables together, capture complex interactions, and assess long-range temporal and anisotropic spatial correlations.

Canonical ML treats the detection problem as a one-class classification problem or an outlier detection problem. Many methods are thus applied²³ and available in software packages². Recent advances include deep learning for segmentation and detection of tropical cyclones and atmospheric rivers (ARs) in high-resolution climate model output²⁴ and semi-supervised localization of extremes²⁵. Alternatively, reconstruction-based models (e.g., with autoencoders) are optimized to accurately reconstruct normal data instances, and thus extremes are associated with large reconstruction errors²⁶. Finally, probabilistic approaches try to identify extremes by estimating the data probability density function (PDF) or specific quantiles thereof. Standard extreme value theory (EVT) approaches, however, struggle with short time-series data, nonlinearities, and nonstationary processes. Alternative probabilistic (ML) models have relied on surrogate meteorological and hydrological seasonal re-forecasts²⁷, Gaussian processes and nonlinear dependence measures²⁸ and multivariate Gaussianization²⁹. Probabilistic approaches in ML also allow to derive confidence intervals for the predictions or even be optimized to certain (high, extreme) levels of interest using quantile regression, tail calibration approaches³⁰, or multivariate EVT³¹.

Prediction. Designing predictive systems that accurately model extreme events is essential to anticipating the effects of future extreme

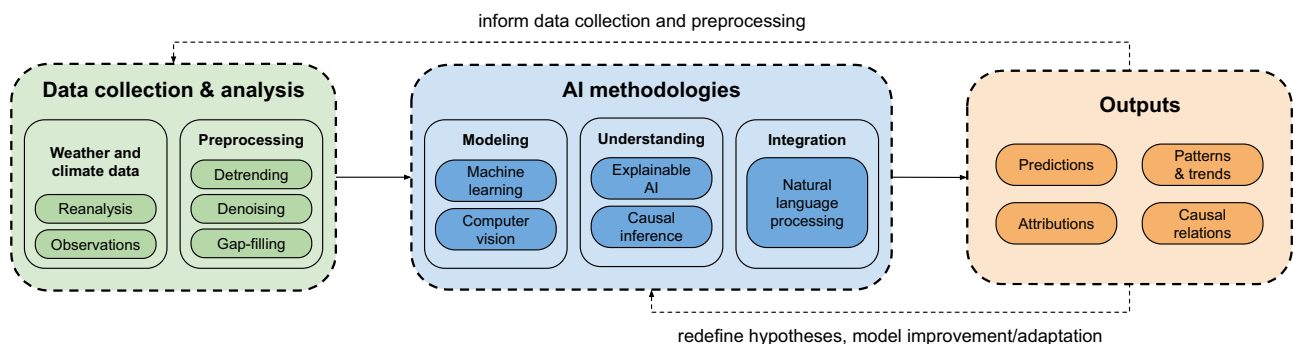


Fig. 1 | A general Artificial Intelligence (AI) driven extreme event analysis pipeline. Different components in modeling and understanding extreme events using AI methodologies are interconnected, highlighting the flow from data collection and analysis to actionable insights/outputs. Note the feedback loops where AI does not only produce some relevant outputs and products from data

(predictions, patterns and trends, climate attribution, and causal relations) but also may help suggest redefining the hypotheses or the improvement and adaptation in methodologies to overcome identified challenges, as well as inform data collection and preprocessing.

Table 1 | Comprehensive categorization of extreme events analysis

Aspect	Task	Standard methods	Current AI methods
Data	Data collection, harmonization, preprocessing Gather, standardize, and clean relevant multi-source data for the down-stream task	Manual normalization and detrending; statistical methods for denoising, gap-filling, and data fusion	Automated ML/DL-based pipelines for denoising, gap-filling, and data fusion; anomaly detection
	Feature selection and extraction Find informative low-dimensional representations of the data	Statistical tests and dimensionality reduction, manual feature selection, and feature engineering	Automated feature selection via feature importance; deep feature extraction and embedded methods; causal feature selection and representation learning
	Simulation of extremes Simulate extremes under varying scenarios with correct frequency and amplitude	Physical models, Monte Carlo simulations, extreme value theory	ML-based, physics-informed, and hybrid generative models
Modeling	Detection Detect and identify extreme events geographically and over time	Threshold- or percentile-based methods, extreme value theory	Outlier detection, one-class classification, reconstruction-based models, probabilistic methods
	Prediction Make predictions on potential future extreme events	Numerical weather prediction models, expert judgment, heuristics, time-series analysis	Deterministic and probabilistic ML/DL forecasting models, recurrent neural networks
	Impact assessment Estimate the effects of extreme events on society, the economy, and the natural environment	Expert judgment, heuristics, field data collection, statistical impact analysis	Simulation-based ML/DL models, deterministic and probabilistic ML/DL methods for predictive impact modeling
Understanding and trustworthiness	Explainability Make the rationale of a black-box ML method understandable by humans	Transparent and interpretable by design models (e.g., linear approximations, decision trees)	Distillation methods, feature attribution methods, attention mechanisms
	Causality Identify and analyze factors directly leading to the occurrence of extreme events	Statistical and correlation-based analysis, simulation-based, and experimental studies	Causal inference, counterfactual explanations
	Extreme event attribution Quantify influence of anthropogenic forcings on extreme events	Statistical event attribution, storyline approaches, expert judgment, and elicitation	Weather and climate emulation models, ensemble techniques
	Uncertainty quantification Measure the reliability of predictions, identify and quantify sources of uncertainty	Confidence intervals, sensitivity analysis, parametric distributions	Bayesian methods, ensemble methods
Last mile	Operationalization Integrate trustworthy predictive models and real-time data analysis into early warning systems	Manual integration into decision systems	Automated integration using APIs and ML-driven decision support, cloud-based deployment
	Communication of risk and ethical aspects Convey information on extreme risks, ensuring ethical transparency, community preparedness and response	Adherence to ethical guidelines and standards, risk reports	Ethical and fairness-aware AI models, causality for quantification of biases
	Policy and Decision-making Manage and mitigate impacts of extreme events, ensuring resilience and safety	Risk assessment and management, cost-benefit analysis, multi-criteria decision analysis, adaptive management, and community-based participatory approaches	Time series analysis, ensemble methods, anomaly detection, clustering, reinforcement learning

A taxonomical organization of the analysis of extreme events in four main blocks (data, modeling, understanding, and trustworthiness, and the last mile on communication and deployment) regarding their task, standard (e.g., dimensionality reduction and feature selection, correlation analysis, statistical event attribution, risk reports) and advanced AI methods (e.g., deep representation learning, distillation, causal inference, and counterfactual analysis, automated AI pipelines, reinforcement learning).

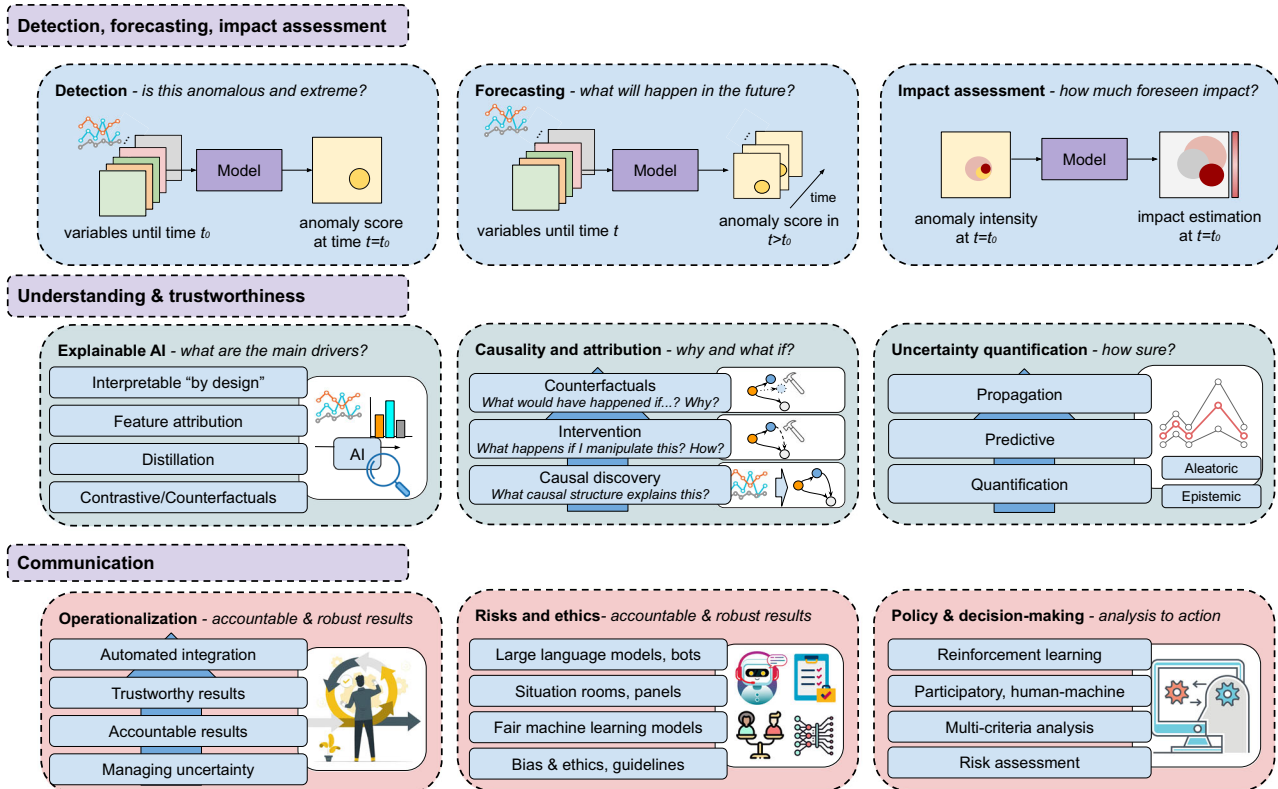


Fig. 2 | Components in an Artificial Intelligence (AI) pipeline for extreme events. AI mainly exploits spatio-temporal Earth observation, reanalysis, and climate data to answer “what”-questions (top row): detection of events, prediction, and impact assessment. AI can also be used for understanding events and thus answer “what if,” “why,” and “how sure” questions (middle row) and makes use of explainable AI (XAI) to identify relevant drivers of events, causality to

understanding the system, estimate causal effects and impacts, and imagine counterfactual scenarios for attribution and uncertainty estimation to quantify trust and robustness for decision-making. Communicating extreme events and their impacts can benefit from statistical/machine learning (bottom row) by improving operationalization, ensuring fair and equitable narratives, and integrating language models in situation rooms for enhanced decision-making.

events and providing critical information for decision-makers to prevent damages and losses. Spatial and temporal predictions aim to provide a quantitative estimate of the future value of the Earth’s state (Table 1).

Many ML algorithms have been proposed for deterministic extreme event prediction, but most are limited to small regions and specific use cases, though. Prediction can be performed using climate variables alone³² or in combination with satellite imagery³³. A common approach is directly estimating an indicator that defines the extreme event, e.g., flood risk maps³⁴ or drought indices³⁵. The estimated variables can cover different lead times depending on the characteristics of the extreme, spanning from short-term prediction to seasonal prediction³⁶. Recently, DL-based prediction techniques have gained popularity for their ability to process large data volumes, capture complex nonlinear relationships, and reduce manual feature engineering. These benefits have led to the creation of global models that generalize across various locations, as seen in flood³⁷ and wildfire³⁸ predictions. Hybrid AI-based techniques integrated within climate models can enhance predictions, such as in drought prediction³⁹ and extreme convective precipitations downscaling⁴⁰. Probabilistic models, unlike deterministic ones, focus on predicting the probability distribution of future variable states. The importance of probabilistic forecasts for extreme heatwaves has been emphasized⁴¹.

Impact assessment. Estimating the effects of extreme events on society, the economy, and the environment is crucial for conveying potential future consequences to the public, policymakers, and across disciplines^{20,42}. Impact assessment involves understanding how a

system reacts to extreme event forcings. Unlike extreme event detection and prediction, the focus here is on impact-related outcomes, such as the number of injuries, households affected, or crop losses.

In recent years, there has been an increasing interest in predicting vegetation state using ML⁴, as a way to estimate the impact of climatic extremes on the evolution of the vegetation state variable^{43,44}. Recent advances methods have used echo state networks⁴⁵, Convolutional Long Short-Term Memory-based (ConvLSTM-based)^{33,46}, and transformers⁴⁷ using high-resolution remote sensing and climate data.

Another way to address impact assessment with ML is to analyze changes in the PDF over time. This approach allows a quantification of the impact of different events, which can be used to improve our understanding of the drivers of vulnerability, such as population displacement⁴⁸. Alternatively, the impact of extreme events can be detected by analyzing the news coverage based on natural language processing (NLP) and, more recently, on LLMs⁴⁹.

Extreme events understanding and trustworthiness

Yet, all previous approaches focus on the “what,” “when,” and “where” questions, not the “why,” “what if,” and “how confident” ones. The latter questions are essential for achieving trustworthy ML⁵⁰, as high-stakes decisions impacting public safety, health, infrastructure, and resource allocation depend on it. Disciplines like XAI and uncertainty quantification (UQ) offer methods to make AI more reliable and trustworthy (see Fig. 2 [middle row]). These approaches not only help us interpret AI model predictions but also enhance our understanding of extreme events themselves. Techniques such as causal inference

and extreme event attribution further complement XAI and UQ by understanding the mechanisms behind these events, crucial for improving AI models and gaining trust in the decision-making process.

Explainable AI. Many models are transparent and interpretable by design, like linear models or decision trees. Still, they may not perform well on complex problems and are rarely employed in extreme events (cf. Table 1). XAI aims to unveil the decision-making process of AI models. XAI also facilitates debugging, improving models, and gathering scientific insight by revealing the model functioning, learned relationships, and biases. The most commonly used XAI approaches rely on model-agnostic *distillation* or *feature attribution* methods^{51,52}, cf. Fig. 2. Distillation methods, such as SHapley Additive exPlanations and Local interpretable model-agnostic explanations, create surrogate models and have been widely used in geosciences and climate sciences^{48,53}. Feature attribution methods, like Partial Dependence Plot (PDP) or Gradient-based Class Activation Map, highlight important features by perturbing the inputs or using backpropagation⁵⁴. Recent approaches explain DL models using attention in drought prediction⁵⁵ and prototypes for explaining event localization⁵⁶. XAI has also been used to evaluate climate predictions^{13,57}, and thus offers an agnostic, data-driven approach for model-data intercomparison. Nevertheless, most XAI methods, particularly post-hoc approaches, approximate the underlying model, and the lack of a clear definition for a “sufficient” explanation complicates their evaluation. Further discussion on XAI’s limitations is available in “Model challenges”.

Causality and attribution. Causality or causal inference aims to uncover the underlying relationships between variables to determine not only what influences an outcome but also why and to what extent. This involves two main tasks: *causal discovery*, which identifies the structure of causal dependencies, and *causal effect estimation*, which quantifies the impact of specific causes on outcomes^{58–60}. Knowing causal relationships allows us to answer *what if* questions like, “What would change with controlled burns or flood barriers?” (interventions) or “What if there were no anthropogenic emissions?” (counterfactuals). For extreme events, most widespread causal inference methods^{60,61} cannot be applied directly, as traditional assumptions of normality and mild outliers do not hold, and dependences may differ between extremes and the tail of the distribution (Table 1). This has motivated recent works that deal with causal discovery for extreme event analysis, providing different frameworks for understanding the dependencies and causal structures in extreme values⁶², with applications on, e.g., rivers discharge⁶³. Another relevant line of work focuses on answering counterfactual questions for climate extreme events, which links to extreme event attribution^{14,64,65}.

Extreme event attribution (EEA) quantifies the influence of anthropogenic forcings (such as greenhouse gas emissions) on the likelihood of extreme climate events and thus aims to answer an inherently causal question. The methods use numerical simulations with General Circulation Models (GCM) to compare their probabilities under observed conditions (the factual world) and a hypothetical scenario without human emissions (the counterfactual world)⁶⁶ (cf. Table 1); or statistical methods applied to observational data. Two main viewpoints exist: probabilistic EEA employs quantitative statistical methods to estimate this likelihood (e.g., the EE is 600 times more likely due to human emissions), while storyline approaches simulate the evolution of the EE under different forcings to gather a process-based attribution statement (e.g., 50% of the magnitude of an EE is explained by natural variability)⁶⁷. Ideally, both methods should be combined to provide a complete understanding. When it comes to the attribution of long-term trends, ensembles of neural networks have been developed to emulate GCMs⁶⁸. Neural network models now successfully predict the year based on the global annual temperature or precipitation field under present conditions, thus effectively

detecting and/or attributing long-term trends⁶⁹, leading to some interesting XAI attribution patterns (not to be confused with EEA, see “Explainable AI”)⁷⁰. In this context, the fingerprint of climate change has been identified in daily global patterns of temperature and humidity⁷¹, and precipitation⁷². More recent work uses these statistical and ML approaches to detect trends in extreme events such as heavy precipitation^{11,73}. Yet, literature on AI uses for formal climate attribution of extreme events is still scarce. However, individual studies have now started developing climate change counterfactuals for heat waves⁶⁵, while AI, as of yet, is more widely used for the detection, prediction, and driver identification of heat waves (see case study on heat waves). More broadly, recent advances in weather and climate emulation with AI^{3,74} suggest that climate emulators may play a significant role in future climate attribution studies.

Uncertainty quantification. Even equipped with explanatory and causal methods, it is still crucial to assess how confident AI model decisions are, as inaccurate warnings or decisions could impact safety and resources^{8,75}. Understanding the sources of uncertainty is important to inform and disentangle the inherent (aleatoric) uncertainty of the weather phenomenon from the lack of knowledge in the model (epistemic) uncertainty. One can reduce the latter with more data and additional assumptions, but not the former (cf. Table 1). DL approaches with UQ have shown promise for extreme events recently^{41,76}.

The Last Mile: operationalization, communication, ethics and decision-making

The previous components (methods and techniques) in the AI pipeline need to be operationalized, be robust, accountable and fair, so they ultimately serve the purpose of evidence-based policy making (see Fig. 2 [bottom row]).

Operationalization. Operationalization strictly needs the previous layer in place (XAI, causality, UQ) for improved and accountable explanations of the predictions in early warning systems (EWSs), and for enhanced disaster risk management. Calibration of ML models⁷⁷, especially when coupled with uncertainty estimation, aligns predicted probabilities to reflect the real-world likelihood of extreme outcomes, enhancing reliability and interpretability essential for informed decision-making. Poor calibration can lead to misguided decisions, such as overestimating or underestimating the likelihood of critical events.

Communication of risk and ethical aspects. Global initiatives like the Common Alerting Protocol have standardized warning data, enabling timely alerts. However, these systems’ effectiveness depends on inclusivity and adaptability to diverse community needs^{78,79}. Achieving this globally remains challenging, as centralized models are easier and cheaper but lack local sensitivity. AI can address this by identifying at-risk populations, optimizing message clarity, selecting effective communication channels, and overcoming the limitations of traditional, one-size-fits-all methods at scale.

Yet, history shows that even actionable forecasts can fail if not communicated properly. For instance, despite predictions of the Mediterranean storm Daniel’s landfall four days in advance, the lack of effective communication contributed to the tragic outcome in Libya, with severe casualties and displacement⁸⁰. Even in the more developed countries of Germany (2021) and Spain (2024), the devastating effects of floods, with more than 200 fatalities each, were related to ineffective warnings^{81,82}. This highlights the critical need for robust EWSs that predict events and effectively communicate risks to ensure community preparedness and response (Table 1)⁷⁸. The challenge of false alarms in the context of EWS is a significant challenge, as they can lead to “warning fatigue”, where the public becomes desensitized to alerts and may ignore crucial warnings during actual emergencies^{79,83,84}.

Addressing this issue requires improving the accuracy and trust in predictive models, refining communication strategies, and engaging the community⁸⁵. AI enhances risk communication by enabling message personalization and improving clarity, especially in real-time systems. Nevertheless, even if AI can help manage and evaluate risk in early warning systems, we think that moving from the warning to the decision, especially when dealing with highly uncertain scenarios, will require communication science, behavioral psychology, and brave decision-makers.

Ethics in AI models and data. The governance of AI ethics calls for systems to respect human dignity, ensure security, and support democratic values⁸⁶. The deployment of AI to help manage extremes involves several fundamental principles: ensuring fairness, maintaining privacy, and achieving transparency^{87,88}. In this context, the rise of LLMs heightens ethical risks like outdated or inaccurate data, bias, errors, and misinformation, often obscured by their vast training on web data. Additionally, current AI models (like generative AI and LLMs) often draw on large datasets, so there is a risk of strong biases if these datasets are unrepresentative⁵². Spatial sampling and analysis are vital to collecting geographically and environmentally representative, fair, and unbiased data. AI enhances EWSs by enabling the rapid dissemination of personalized alerts tailored to specific locations and individual risk factors like proximity to flood plains or wildfire zones, ensuring warnings are understandable and relevant to all.

Achieving the inclusion of affected communities, especially in the Global South, still remains a significant challenge, as centralized, 'one-size-fits-all' models are often simpler and cheaper to implement but less effective at addressing local nuances. Designing AI systems, building on large-language models fine-tuned to local communities with users in the loop, offers excellent opportunities to overcome "one-size-fits-all" while being efficient.

Policy and decision-making. Even if AI-assisted, human operators are in charge of implementing the final decision. This implies supporting end-users in making the most out of the generated information⁶. This operational value is often problem-specific and depends upon the system's dominant dynamics and the socio-economic sectors considered. Quantifying this value, however, might not be straightforward as a more accurate prediction does not necessarily imply a better decision by the end users⁸⁹. When multiple forecasts from different systems are available, users should address a number of problems where AI can be helpful, including the selection of the forecast product, the lead time, the variable aggregation, the bias correction, and how to cope with the forecast uncertainty.

Integrating the pipeline in Fig. 1 with impact models that leverage Reinforcement Learning algorithms to simulate optimal decision-making can help quantify how AI-enhanced information about extreme events translates into better decisions⁹⁰. The co-design of impact models via participatory processes, including end users in the loop, further strengthens the overall model-based investigation by better capturing end users' requirements, expectations and concerns. End users should understand both the benefits of AI-enhanced information for improved decision-making and the risks of misinformation leading to poor choices.

Data, model, and integration challenges

Extreme event analysis faces many important challenges related to the data and model characteristics, but also to the integration of AI in decision pipelines (see Table 2).

Data challenges

A major challenge is the lack of sufficient data with expert annotations, which are essential for training and evaluating AI models. Given their rarity, extreme events can be overlooked during the data

preprocessing steps to eliminate noise, gaps, biases, and inconsistencies^{91,92}. Additionally, AI struggles with integrating and extracting relevant information across various data sources and scales, which can complicate feature extraction and selection⁵². Future AI development needs to focus on deriving robust features (or representations) that effectively capture the distinct characteristics of extreme events (cf. Table 2).

AI models are increasingly used to enhance parameterizations of subgrid-scale processes in Earth system models, addressing gaps in traditional methods. However, a key challenge is their numerical instability over extended time horizons, which may produce unrealistic scenarios when simulating extremes due to insufficient training data⁹³. The quality of observational data also challenges AI methods used for data assimilation. Recently, hybrid models combining domain-driven with data-driven models promises more robust and trustworthy AI models^{92,94}. Moreover, AI-based methods have improved error characterization, enhancing UQ⁹⁵. Generative models are also used to sample ensemble members more effectively, offering better representations of the system's state⁹⁶ (cf. Table 2).

Model challenges

The lack of a clear statistical definition of extreme events and the mechanisms responsible for their occurrence hamper model development and adoption. For detection, extremes usually constitute not pointwise but complex contextual, group, or conditional anomalies, whose source (changing process, parents, distribution) is often unknown⁵². This results in challenges, such as capturing subtle (new) patterns, setting adaptive thresholds, or integrating data across distant points in space and time (cf. Table 2). For prediction and impact assessment, AI models are sensitive to initial conditions and may not capture long-term dependencies⁹¹. Moreover, data might not reveal the dynamics of extremes, as there might be changes to unseen dynamics²³ and stationarity may not hold in general. Non-stationarity and distribution changes make prediction challenging because shifting baselines and evolving data distributions hinder models from generalizing beyond historical data. Also, relationships between variables may change over time, meaning predictions often lack robustness under unseen conditions. Using hybrid models that combine data, domain knowledge, and ML could allow insights into the mechanisms that trigger extreme events^{92,94,97,98}.

The complexity of extremes also makes attribution, causal discovery, and explainability particularly challenging. XAI can only reveal correlations the model learned and has no information about the causal structure. This could lead XAI to exacerbate model biases or spurious correlations. Indeed, different XAI methods can produce very different explanations, whose suitability for different models should be quantitatively evaluated^{13,99}. XAI explanations are also challenging to interpret, often highlighting complex relationships that require expert knowledge¹⁰⁰. Causality is not error-free either, as a wrong assumption about where the extreme comes from could lead to wrong causal graphs, conclusions, and decisions⁵⁸. Finally, challenges in UQ include differentiating and quantifying aleatoric and epistemic uncertainties, which is complicated by the models' over-parameterization and their lack of robust probabilistic foundations (cf. Table 2).

Integration challenges

ML models are typically and necessarily trained on high-quality, well-curated datasets, such as Copernicus ERA 5 reanalysis or cloud-free satellite imagery, which often do not reflect the error-prone meteorological forecasts and cloudy conditions encountered in real-world situations (Table 2). Recent ML approaches propose deploying domain adaptation and transfer learning techniques that find robust, physics-aware, and invariant feature representations to alleviate the problem of dealing with distribution shifts^{92,98,101}.

Table 2 | Challenges, risks and needs in AI for extreme events

Aspect	Challenges	Risks	Future AI
Data	<ul style="list-style-type: none"> - Manage inconsistencies and biases in data - Handle multimodal data from diverse sources - Accommodate variations in data resolutions - Handle sparse occurrences of extreme events - Adapt to evolving datasets 	<ul style="list-style-type: none"> - Lack of sufficient data with expert annotations - Low number of samples for the anomalous case - Difficulty in defining what constitutes an extreme event - Lose critical extreme values when preprocessing - Discriminative information may be lost - Computationally expensive simulations 	<ul style="list-style-type: none"> - Transfer learning - Class-imbalanced and low-shot learning - Long-tail learning - Online and continual ML - Generative and foundation models - Attention mechanisms and transformers - Causal representation learning - Physics-based and hybrid ML models
Model	<ul style="list-style-type: none"> - Manage complex contextual anomalies - Integrate data across distant space-time points - Capture long-term dependencies - Capture subtle (new) patterns while minimizing false positives and negatives - Set adaptive thresholds 	<ul style="list-style-type: none"> - Unknown sources of anomalies - Sensitivity of AI models to initial conditions - Data might not reveal the dynamics of extremes - Changes to unseen dynamics - Stationarity may not hold - Models with insufficiently heavy tails 	<ul style="list-style-type: none"> - Semi-, self-, unsupervised learning - Multimodal learning - Graph neural networks - Physics-based and hybrid ML models - ML/DL with forecasts and simulations as input features - Reinforcement learning
Understanding and trustworthiness	<ul style="list-style-type: none"> - Attribution of weather and climate extremes - Explanations for out-of-distribution samples - Causal dependence given extremes - Uncertainty calibration in the tails of the PDF 	<ul style="list-style-type: none"> - Wrong assumption on where the anomaly comes leads to wrong causal graphs - Complex xAI methods, requiring additional AI models - Different xAI approaches provide different explanations - Difficulty in finding explanations and causal relations due to complex data relationships - UQ generally difficult in DL due to high dimensionality 	<ul style="list-style-type: none"> - Benchmarking and new evaluation methodologies - Attention mechanisms - Prototype-based models - Causal inference - Variational inference - Gaussian processes
Integration	<ul style="list-style-type: none"> - Scale to large datasets and real scenarios - Generalization and transferability - Bias, transparency and fairness - Human-AI interaction and decision support - Communicate and manage uncertainties 	<ul style="list-style-type: none"> - Data quality, availability, complexity, and interpretability - Lack of domain expertise and collaboration - Outputs difficult for non-experts to interpret or trust - Missing uncertainty quantification for models and graphs - Unintentionally perpetuate biases in decision-making - Resistance to use AI-driven decision support systems 	<ul style="list-style-type: none"> - Domain adaptation - Replicability and validity, human-in-the-loop - Libraries and improved implementations - Distributed solutions and federated learning - Model calibration and uncertainty quantification - Large language models - AI for perception and reasoning

Categorization of the main challenges and risks for extreme event analysis using AI, as well as the needed advances for a robust and trustworthy AI for the presented blocks in "introduction" and Table 1, related to data, model, and integration challenges.

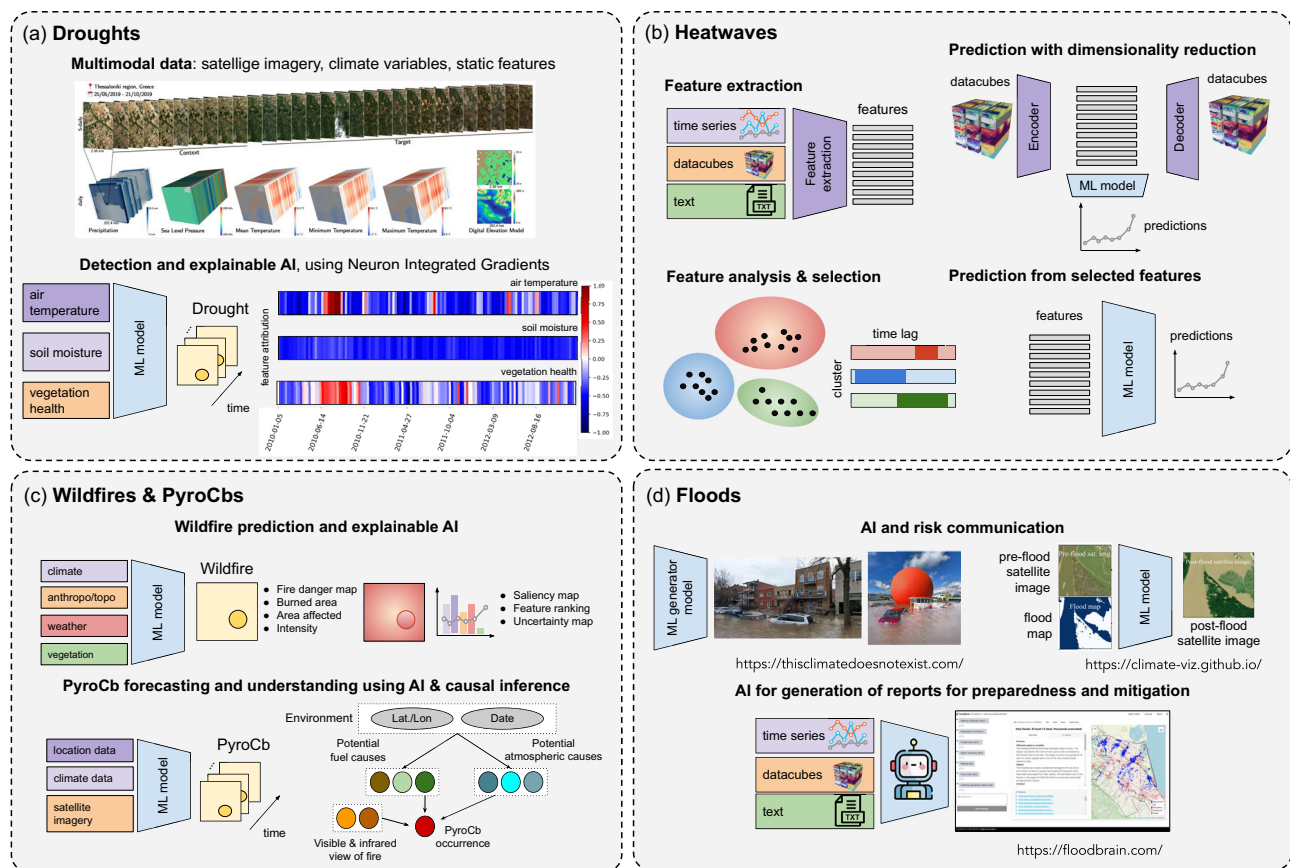


Fig. 3 | Summary of case studies using Artificial Intelligence (AI) to manage extreme events. Four case studies (drought, heatwaves, wildfires, and floods) are showcased where AI enables detection, forecasting, impact assessment, explanation, understanding, and communication of risk, providing a comprehensive solution for disaster management. **a Droughts.** Top: AI leverages multimodal data to predict Earth's surface dynamics, enhancing forecasts for crop yields, forest health, and drought impacts. Bottom: XAI techniques, like “neuron integrated gradients,” elucidate the key factors driving severe drought conditions, highlighting variable interactions over time. **b Heatwaves.** Top: Variables of interest are extracted from heterogeneous data sources (images, time series, text) and potentially aggregated over space and/or time. Bottom Left: Relevant features can be extracted from the data using clustering techniques, for example. Right: Heatwave prediction can be done by combining dimensionality reduction tools or directly from the selected features. **c Wildfires.** Top: AI enhances understanding and prediction of wildfire dynamics, particularly for mega-fires intensified by global warming, by analyzing

extensive datasets and differentiating fire types with XAI. Bottom: AI combined with causal inference aims to better detect and understand pyrocumulonimbus clouds, intense storm systems generated by large wildfires that complicate fire behavior prediction. **d Floods.** AI transforms flood risk communication by using realistic 3D visualizations and animations to depict rising water levels' impact on communities and infrastructure, making the information more relatable (thisclimatedoesnotexist.com). AI-driven platforms analyze vast amounts of data from weather forecasts, river levels, and historical flood patterns to predict future events accurately, integrating this information with digital maps and urban models to identify high-risk areas ([climate-viz.github.io.com](https://climate-viz.github.io/)). This approach enhances flood risk management by allowing targeted, personalized communication, enabling residents to receive specific alerts and visualize potential impacts on their homes. AI also supports the generation of detailed flood reports from various sources, enhancing preparedness and mitigation efforts (floodbrain.com).

Applying domain adaptation strategies or leveraging invariant features could align model performance from training phases to operational conditions. Besides, leveraging proprietary and trusted geospatial data from operational stakeholders, such as detailed forest fuel maps and elevation models, allows fine-tuning these models to enhance detection and forecasting accuracy and enable finer spatial and temporal resolution in output products.

Case studies

We showcase the research in AI for extreme events through four case studies, each focusing on a distinct hazard: droughts, heatwaves, wildfires, and floods. In each case, we highlight the current research landscape and what novel, pressing questions AI can help address (Fig. 3).

Droughts: from detection to impact assessment

Droughts are among the costliest natural hazards, having destructive effects on the ecological environment, agricultural production, and

socioeconomic conditions. While “conventional droughts” are typically seasonal or recurring events that follow predictable patterns based on regional climate, “extreme droughts” are severe, prolonged, and increasingly influenced by climate change, often marked by multi-year impacts, record low water levels, and more pronounced economic and ecological damage. Their detection has been traditionally approached with heuristic thresholds and simple parametric models on related variables (e.g., soil moisture, vegetation indices, precipitation, temperature, etc.). Modern statistical and ML methods enhance drought estimation by leveraging vast Earth observation data from satellites and dense networks of measurement stations¹⁰².

Drought definitions presented in the literature are usually driven by a categorization into the main factors or impacts (meteorological, hydrological, agricultural, operational, or socioeconomic), but is at the same time subjective¹⁰³. These types are not independent but refer to complementary physical, chemical, and biological processes involved in droughts. Although empirical multivariate drought indicators have been proposed to account for these dependencies, the complexity of

the combination of drought processes makes their detection, prediction, and characterization (severity, duration, start-end, etc.) a burden to researchers and policymakers, which can benefit from data-driven AI approaches.

Thanks to their ability to recognize abnormal patterns in high-dimensional (e.g., multivariate spatio-temporal data) spaces, ML models can help provide an alternative, agnostic, data-driven definition of extreme drought events. ML approaches have been successfully applied to drought prediction, from support vector machines, decision trees, and random forests to more advanced multivariate density estimation approaches^{23,29,98}. However, DL algorithms typically improve results by exploiting spatial and temporal data relations more efficiently. Neural networks have been used for drought monitoring in a supervised way, mainly using multilayer, convolutional, and/or recurrent neural networks². However, it is acknowledged that drought detection is very challenging due to scarce, unreliable, or even non-existent labels, making it well-suited for unsupervised and semi-supervised approaches¹⁸.

Recent studies show DL's ability to integrate multimodal data sources (e.g., satellite imagery, mesoscale climate variables, and static features) along with physical knowledge in hybrid models, such as domain-informed variational autoencoders (VAEs) that combine traditional drought indicators with ancillary climate data⁹⁸. These approaches help pinpoint critical regions and key drought drivers¹⁰⁴ (Fig. 3a, bottom). AI-based forecasting can, in turn, answer questions about drought's impact on crop yield and forest health⁴⁷ (Fig. 3a top).

Heatwaves: from prediction to understanding drivers

In the context of climate change, heat extremes are becoming more frequent and intense¹⁰⁵, but still, regional trends are uncertain¹⁰⁶. For example, a notable warming hotspot in Western Europe was identified with heat extremes increasing faster than predicted by the current state-of-the-art climate models¹⁰⁷. Hence, several drivers of heat extremes remain uncertain, specifically long-term changes in atmospheric circulation and land-atmosphere interactions that may amplify or enhance specific events¹⁰⁵. Yet, understanding the proximate physical drivers of heat extremes is crucial for accurate prediction, characterization, and understanding of these events. In that sense, AI models have shown impressive results to learn to detect and predict heat extremes on different spatial and temporal scales, some lead times and aggregations using deep learning^{12,108,109}, causal-informed ridge regression³², or hybrid models^{110,111}, among others. These approaches often require dimensionality reduction tools to select predictors from high-dimensional climate data (Fig. 3c). Expressive lower-dimensional feature representations have been extracted from high-dimensional data using VAEs, which can be used to improve the prediction of heatwaves⁴¹ as well as improve physical understanding of heatwaves¹¹².

New methods like the EarthFormer explore transformer networks specifically for predicting temperature anomalies, combining encoder/decoder structures with spatial attention layers¹¹³.

Recent developments show that also for heatwaves, AI is pushing the limits of what is possible and helps us answer new questions. ML models can be designed for extreme event modeling and sampling. In fact, AI models may struggle with modeling the tails of a distribution accurately¹¹⁴. Great attention has been devoted to designing AI algorithms that focus on this challenge, from designing custom loss functions¹¹⁵ to extensions of quantile regression approaches^{30,116}.

XAI and causal inference have helped understand the physical drivers behind heat extremes^{32,110}. Beyond this, attribution studies for extreme heat have increasingly used AI techniques^{65,117}. Identifying circulation-induced vs. thermodynamical changes becomes crucial when it comes to heatwaves. Several statistical and ML methods are currently in use for this purpose¹¹⁸. ML can further address key questions in attribution, such as the role of dynamical changes in observed

trends, helping to reconcile circulation trends in climate models with reanalysis data^{107,118}.

Wildfires: forecasting risk and understanding processes

Most wildfire damage stems from a few extreme events. Predicting these events is vital for effective fire management and ecosystem conservation. Climate change is expected to increase the frequency, size, and severity of extreme wildfires, exacerbating fire weather conditions¹⁹. Traditional firefighting approaches are increasingly inadequate, with many extreme fires burning until naturally extinguished¹²⁰. The complexity of nonlinear interactions among fire drivers across various scales hinders the predictability of fire behavior. Therefore, enhancing models to better understand and predict conditions leading to large, uncontrollable fires is crucial.

AI can be used to predict the extent and burnt area of wildfires but not the ignition event itself, which is often driven by unpredictable factors like dry lightning (eventually due to atmospheric instability) or other less sampled conditions (e.g., sudden wind shifts, localized fuel conditions, human activities or microscale topography). DL methods can associate weather predictions, satellite observations, and burned area datasets to model wildfires, achieving better predictability than conventional methods. DL has been used successfully to forecast wildfire danger³⁶ and for susceptibility mapping¹²¹. In the short term, wildfires are driven by the daily weather variations but also by the cumulative effects of vegetation and drought. On sub-seasonal to seasonal scales, where weather forecasts are less reliable, wildfires are modulated by Earth system large-scale processes, such as teleconnections. Early work has shown that DL models can leverage information from teleconnections to improve long-term wildfire forecasting¹²².

Beyond predicting wildfire risk, AI and causality aid in understanding key wildfire-related processes. For instance, it is very important to understand the reasons ('the conditions leading to wildfires') behind the predictions to support decision-making and fire management. In this context, XAI can aid in identifying what drives wildfires, supporting, for example, the differentiation and management of various fire types, such as wind-driven and drought-driven fires³⁶. Recent works have advanced in hybrid wildfire spread models¹²³ that also help understanding processes of ignition and spread with deep learning and process understanding jointly.

Due to global warming¹²⁰, we expect an increase in the frequency of unstable atmospheric conditions in the coming years. These conditions can lead to the formation of Pyrocumulonimbus clouds (pyroCb)s—storm clouds that generate their own weather fronts—which can make wildfire behavior unpredictable¹²⁴. Despite the risk posed by pyroCb, the conditions leading to their occurrence and evolution are still poorly understood, and their causal mechanisms are uncertain. In combination with causal inference, AI can advance the detection, forecasting, and understanding of the drivers of pyroCb events¹²⁵.

Floods: from modeling to communication of risk

Studying floods is crucial as they are the most frequent and costly natural disasters, affecting millions annually and causing over \$40 billion in damages worldwide each year¹²⁶. Developing novel methods for flood detection enhances EWSs, reducing fatalities by up to 40%, while effective risk communication ensures better preparedness and response, potentially saving thousands of lives.

In July 2021, intense rainfall in Germany, Belgium, France, and the UK triggered severe flash floods, particularly devastating the Ahr River region in Germany, claiming nearly 200 lives^{127,128}. Alerts were issued by the European Flood Awareness System (EFAS) and disseminated through an international warning system. However, the floods exposed significant flaws in these systems (Fig. 3d). Damage to river monitoring infrastructure impeded data accuracy, and although meteorological forecasts predicted the rainfall two days in advance, flood predictions

for smaller basins lacked precision and failed to consider debris flow and morphodynamics, leading to underestimations of flood severity.

AI offers promising avenues for enhancing flood management systems. Advanced global meteorological forecasting models powered by AI can rapidly process large data ensembles, providing more accurate probabilistic estimates even during extreme weather events. Recent hybrid models such as physics-guided deep learning for rainfall-runoff modeling that consider extreme events¹²⁹ and hybrid DL of the global hydrological cycle¹³⁰ are advancing the field with improved consistency and forecasting capabilities. Furthermore, AI techniques such as ML-accelerated computational fluid dynamics could address the computational challenges in hydro-morphodynamic modeling, allowing for more precise predictions of stream flow and flood levels, particularly in ungauged basins¹³¹.

Additionally, AI can aid in calibrating non-contact video gauges, potentially more robust than traditional methods, as they are not directly affected by the water. AI can also guide forensic analysis to evaluate exposure and vulnerability, employing multi-modal approaches to refine geospatial models at a local scale. However, the effectiveness of these AI-driven models depends on detailed knowledge of the local terrain, including potential bottlenecks like bridges and channels, and accurate data about societal vulnerability to floods⁴⁸.

Yet, AI can also transform how warnings are issued to improve communication and response strategies. For instance, AI-generated maps and photorealistic visualizations based on digital elevation models can depict expected inundation areas and damages¹³². Moreover, AI can generate easily understandable, language-based warnings, both written and auditory, tailored for diverse populations, including the visually impaired. An LLM-based chatbot feature could enhance interactivity, providing real-time, personalized responses to emergency inquiries (see examples in Fig. 3d). AI can improve risk management in EWSs, yet turning alerts into actions in uncertain scenarios requires effective communication^{81,82}.

Conclusions and outlook

This review *highlights* the significant potential of AI for analyzing and modeling extreme events while also detailing the main difficulties and prospects associated with this emerging field. Integrating AI into extreme event analysis faces several challenges, including data management issues like handling dynamic datasets, biases, and high dimensionality that complicate feature extraction. AI models also struggle with unclear statistical definitions of what is ‘extreme.’ Furthermore, integrating AI with physical models poses substantial challenges yet offers promising opportunities for enhancing model accuracy and reliability. Trustworthiness concerns arise from the complexity and interpretability of ML models, the difficulty of generalizing across different contexts, and the quantification of uncertainty. Operational challenges include the complexity of AI outputs, which hinder interpretation by non-experts, resistance to AI adoption due to concerns over reliability and fairness, and the need for frameworks that facilitate transparent and ethical integration of AI insights into decision-making processes.

The previous challenges compromise the reproducibility and comparability of ML models in analyzing extreme events. These challenges are exacerbated by data scarcity, lack of transparency in model configurations, and the use of proprietary tools. Additionally, interdisciplinary differences hinder consistency and comparability. Effective solutions demand robust, transparent methodologies, inclusive data-sharing practices, and frameworks that support cross-disciplinary collaboration.

In this review, we overviewed and emphasized the importance of developing operational, explainable, and trustworthy AI systems. Addressing these challenges requires a coordinated effort across disciplines involving AI researchers, environmental and climate scientists, field experts, and policymakers. This collaborative approach is crucial

for advancing AI applications in extreme event analysis and ensuring that these technologies are adapted to real-world needs and constraints. From an operational perspective, adapting AI solutions to real-time data integration, model deployment, and resource allocation highlights the need for systems that can function within the operational frameworks of disaster management and risk mitigation. Besides, methodological improvements are still needed in model evaluation and benchmarking to alleviate issues like overfitting and enhance AI systems’ generalization capabilities.

Looking forward, there are significant areas for further exploration and improvement. These include the development of benchmarks specific to extreme events, enhanced integration of domain knowledge to improve data fusion and model training, and the creation of robust, scalable AI systems capable of adapting to the dynamic nature of extreme events. Recent LLMs harvest vast amounts of domain knowledge embedded in the literature, promising significant advances in communicating risk.

As we advance, the ultimate goal is to harness AI’s potential to substantially benefit society, particularly by enhancing our capacity to manage and respond to extreme events. Through dedicated research and collaborative innovation, AI can become a cornerstone in our strategy to understand and mitigate the impacts of these challenging and elusive phenomena.

References

1. Seneviratne, S.-I. et al. *Weather and Climate Extreme Events in a Changing Climate* (Cambridge University Press, 2021).
2. Gonzalez-Calabuig, M. et al. The AIDE toolbox: AI for disentangling extreme events. *IEEE Geosci. Remote Sens. Mag.* **12**, 1–8 (2024).
3. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1422 (2023). **This study introduces a machine learning model that enhances medium-range global weather forecasts, outperforming traditional methods in accuracy.**
4. Ferchichi, A., Abbes, A. B., Barra, V. & Farah, I. R. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: a systematic literature review. *Ecol. Inform.* **68**, 101552 (2022).
5. Ragone, F. & Bouchet, F. Rare event algorithm study of extreme warm summers and heatwaves over Europe. *Geophys. Res. Lett.* **48**, e2020GL091197 (2021).
6. Yokota, F. & Thompson, K. Value of information analysis in environmental health risk management decisions: past, present, and future. *Risk Anal.* **24**, 635–650 (2004).
7. Salcedo-Sanz, S. et al. Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: a review. *Theor. Appl. Climatol.* **155**, 1–44 (2024).
8. Gawlikowski, J. et al. A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.* **56**, 1513–1589 (2023).
9. Harrington, L. J., Schleussner, C. F. & Otto, F. E. Quantifying uncertainty in aggregated climate change risk assessments. *Nat. Commun.* **12**, 7140 (2021).
10. Stott, P. A. et al. Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev. Clim. Chang.* **7**, 23–41 (2016). **The paper reviews methodologies linking specific extreme weather events to human-induced climate change, highlighting advancements in attribution science.**
11. Madakumbura, G. D., Thackeray, C. W., Norris, J., Goldenson, N. & Hall, A. Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. *Nat. Commun.* **12**, 3944 (2021).
12. Chattopadhyay, A., Nabizadeh, E. & Hassanzadeh, P. Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Model. Earth Syst.* **12**, e2019MS001958 (2020).

13. Bommer, P. L., Kretschmer, M., Hedstrom, A., Bareeva, D. & Hohne, M. M.-C. Finding the right XAI method—a guide for the evaluation and ranking of explainable AI methods in climate science. *Artif. Intell. Earth Syst.* **3**, e230074 (2024).
14. Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P. & Ghil, M. Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Am. Meteorol. Soc.* **97**, 99–110 (2016). **The authors develop a causal counterfactual framework to assess the influence of specific factors on weather and climate events, enhancing attribution accuracy.**
15. Vaghefi, S. A. et al. ChatClimate: grounding conversational AI in climate science. *Commun. Earth Environ.* **4**, 480 (2023).
16. Olivetti, L. & Messori, G. Advances and prospects of deep learning for medium-range extreme weather forecasting. *Geosci. Model Dev.* **17**, 2347–2358 (2024).
17. Materia, S. et al. Artificial intelligence for climate prediction of extremes: state of the art, challenges, and future perspectives. *WIREs Clim. Chang.* **n/a**, e914 (2024).
18. Ruff, L. et al. A unifying review of deep and shallow anomaly detection. *Proc. IEEE* **109**, 756–795 (2021).
19. Han, Z., Zhao, J., Leung, H., Ma, K. F. & Wang, W. A review of deep learning models for time series prediction. *IEEE Sens. J.* **21**, 7833–7848 (2019).
20. Zennaro, F. et al. Exploring machine learning potential for climate change risk assessment. *Earth Sci. Rev.* **220**, 103752 (2021).
21. Jones, R. L., Kharb, A. & Tubeuf, S. The untold story of missing data in disaster research: a systematic review of the empirical literature utilising the emergency events database (EM-DAT). *Environ. Res. Lett.* **18**, 103006 (2023).
22. Mahecha, M. D. et al. Detecting impacts of extreme events with ecological in situ monitoring networks. *Biogeosciences* **14**, 4255–4277 (2017).
23. Flach, M. et al. Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques. *Earth Syst. Dyn.* **8**, 677–696 (2017).
24. Prabhat, et al. ClimateNet: an expert-labelled open dataset and Deep Learning architecture for enabling high-precision analyses of extreme weather. *Geosci. Model Dev. Discuss.* **2020**, 1–28 (2020).
25. Racah, E. et al. Extremeweather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Advances in Neural Information Processing Systems (NeurIPS)* **30** (2017).
26. Guancho García, Y., Shadaydeh, M., Mahecha, M. & Denzler, J. Extreme anomaly event detection in biosphere using linear regression and a spatiotemporal MRF model. *Nat. Hazards* **98**, 849–867 (2018).
27. Klehmet, K. et al. Robustness of hydrometeorological extremes in surrogated seasonal forecasts. *Int. J. Climatol.* **44**, 1725–1738 (2024).
28. Johnson, J. E., Laparra, V., Pérez-Suay, A., Mahecha, M. & Camps-Valls, G. Kernel methods and their derivatives: concept and perspectives for the Earth system sciences. *PLoS ONE* **15**, p.e0235885 (2020).
29. Johnson, J. E., Laparra, V., Piles, M. & Camps-Valls, G. Gaussifying the Earth: multidimensional information measures for Earth data analysis. *IEEE Geosci. Remote Sens. Mag.* **9**, 191–208 (2021).
30. Allen, S., Koh, J., Segers, J. & Ziegel, J. Tail calibration of probabilistic forecasts. *arXiv preprint arXiv:2407.03167* (2024).
31. Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K. & Engelke, S. Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks. *Environ. Data Sci.* **1**, e5 (2022).
32. Vijverberg, S. & Coumou, D. The role of the Pacific Decadal Oscillation and ocean-atmosphere interactions in driving US temperature variability. *npj Clim. Atmos. Sci.* **5**, 18 (2022).
33. Kladny, K.-R., Milanta, M., Mraz, O., Hufkens, K. & Stocker, B. D. Enhanced prediction of vegetation responses to extreme drought using deep learning and Earth observation data. *Ecol. Inform.* **80**, 102474 (2024).
34. Bentivoglio, R., Isufi, E., Jonkman, S. N. & Taormina, R. Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrol. Earth Syst. Sci.* **26**, 4345–4378 (2022).
35. Belayneh, A., Adamowski, J., Khalil, B. & Ozga-Zielinski, B. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J. Hydrol.* **508**, 418–429 (2014).
36. Kondylatos, S. et al. Wildfire danger prediction and understanding with deep learning. *Geophys. Res. Lett.* **49**, e2022GL099368 (2022).
37. Nearing, G. et al. Global prediction of extreme floods in ungauged watersheds. *Nature* **627**, 559–563 (2024).
38. Zhang, G., Wang, M. & Liu, K. Deep neural networks for global wildfire susceptibility modelling. *Ecol. Indic.* **127**, 107735 (2021).
39. Vo, T. Q., Kim, S.-H., Nguyen, D. H. & Bae, D.-H. LSTM-CM: a hybrid approach for natural drought prediction based on deep learning and climate models. *Stoch. Environ. Res. Risk Assess.* **37**, 2035–2051 (2023).
40. Shi, X. Enabling smart dynamical downscaling of extreme precipitation events with machine learning. *Geophys. Res. Lett.* **47**, e2020GL090309 (2020).
41. Miloshevich, G., Cozian, B., Abry, P., Borgnat, P. & Bouchet, F. Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data. *Phys. Rev. Fluids* **8**, 40501 (2023).
42. Callaghan, M. et al. Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat. Clim. Chang.* **11**, 966–972 (2021).
43. Sutanto, S. J. et al. Moving from drought hazard to impact forecasts. *Nat. Commun.* **10**, 4945 (2019).
44. Salakpi, E. E. et al. Forecasting vegetation condition with a Bayesian auto-regressive distributed lags (BARDL) model. *Nat. Hazards Earth Syst. Sci.* **22**, 2703–2723 (2022).
45. Martinuzzi, F. et al. Learning extreme vegetation response to climate drivers with recurrent neural networks. *Nonlinear Process. Geophys.* **31**, 535–557 (2024).
46. Ahmad, R., Yang, B., Ettlin, G., Berger, A. & Rodríguez-Bocca, P. A machine-learning based ConvLSTM architecture for NDVI forecasting. *Int. Trans. Oper. Res.* **30**, 2025–2048 (2023).
47. Benson, V. et al. Forecasting Localized Weather Impacts on Vegetation as Seen from Space with Meteo-Guided Video Prediction. <https://doi.org/10.48550/arXiv.2303.16198> (2024).
48. Ronco, M. et al. Exploring interactions between socioeconomic context and natural hazards on human population displacement. *Nat. Commun.* **14**, 8004 (2023). **This study uses XAI to explain how socioeconomic factors and natural hazards influence human displacement, revealing that vulnerable communities are disproportionately affected.**
49. Sodoge, J., Kuhllicke, C. & de Brito, M. M. Automated spatio-temporal detection of drought impacts from newspaper articles using natural language processing and machine learning. *Weather Clim. Extrem.* **41**, 100574 (2023).
50. Bostrom, A. et al. Trust and trustworthy artificial intelligence: a research agenda for ai in the environmental sciences. *Risk Anal.* **44**, 1498–1513 (2024).
51. Ghaffarian, S., Taghikhah, F. R. & Maier, H. R. Explainable artificial intelligence in disaster risk management: achievements and prospective futures. *Int. J. Disaster Risk Reduct.* **98**, 104123 (2023).
52. Tuia, D. et al. Artificial intelligence to advance earth observation: a review of models, recent trends, and pathways forward. *IEEE Geoscience and Remote Sensing Magazine*, 2–25 (IEEE, 2024). **The**

- paper discusses the potential of artificial intelligence to enhance Earth observation capabilities, emphasizing its role in improving data analysis and interpretation.**
53. Schlund, M. et al. Constraining uncertainty in projected gross primary production with machine learning. *J. Geophys. Res.: Biogeosci.* **125**, e2019JG005619 (2020).
 54. Srinivasan, R., Wang, L. & Bulleid, J. Machine learning-based climate time series anomaly detection using convolutional neural networks. *Weather Clim.* **40**, 16–31 (2020).
 55. Dikshit, A., Pradhan, B., Assiri, M. E., Almazroui, M. & Park, H.-J. Solving transparency in drought forecasting using attention models. *Sci. Total Environ.* **837**, 155856 (2022).
 56. Barnes, E. A., Barnes, R. J., Martin, Z. K. & Rader, J. K. This looks like that there: interpretable neural networks for image tasks when location matters. *Artif. Intell. Earth Syst.* **1**, e220001 (2022).
 57. Mamalakis, A., Barnes, E. A. & Ebert-Uphoff, I. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artif. Intell. Earth Syst.* **1**, e220012 (2022).
 58. Pearl, J. *Causality: Models, Reasoning, and Inference* 2nd edn (MIT Press, 2017).
 59. Peters, J., Janzing, D. & Schölkopf, B. *Elements of Causal Inference—Foundations and Learning Algorithms* Adaptive Computation and Machine Learning Series (MIT Press, 2017).
 60. Camps-Valls, G. et al. Discovering causal relations and equations from data. *Phys. Rep.* **1044**, 1–68 (2023). **This comprehensive review explores methods for uncovering causal relationships and deriving equations from data, with applications across various scientific fields.**
 61. Runge, J., Gerhardus, A., Varando, G., Eyring, V. & Camps-Valls, G. Causal inference for time series. *Nat. Rev. Earth Environ.* **4**, 487–505 (2023).
 62. Gnecco, N., Meinshausen, N., Peters, J. & Engelke, S. Causal discovery in heavy-tailed models. *Ann. Stat.* **49**, 1755–1778 (2019).
 63. Pasche, O. C., Chavez-Demoulin, V. & Davison, A. C. Causal modelling of heavy-tailed variables and confounders with application to river flow. *Extremes* **26**, 573–594 (2021).
 64. Kiriliouk, A. & Naveau, P. Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *Ann. Appl. Stat.* **14**, 1342–1358 (2020).
 65. Trok, J. T., Barnes, E. A., Davenport, F. V. & Diffenbaugh, N. S. Machine learning-based extreme event attribution. *Sci. Adv.* **10**, eadl3242 (2024).
 66. Naveau, P., Hannart, A. & Ribes, A. Statistical methods for extreme event attribution in climate science. *Annu. Rev. Stat. Appl.* **7**, 89–110 (2020).
 67. Otto, F. E. Attribution of extreme events to climate change. *Annu. Rev. Environ. Resour.* **48**, 813–828 (2023). **The article reviews progress in attributing extreme weather events to climate change, highlighting methodological advancements and challenges in the field.**
 68. Pasini, A., Racca, P., Amendola, S., Cartocci, G. & Cassardo, C. Attribution of recent temperature behaviour reassessed by a neural-network method. *Sci. Rep.* **7**, 17681 (2017).
 69. Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C. & Anderson, D. Viewing forced climate patterns through an AI lens. *Geophys. Res. Lett.* **46**, 13389–13398 (2019).
 70. Barnes, E. A. et al. Indicator patterns of forced change learned by an artificial neural network. *J. Adv. Model. Earth Syst.* **12**, e2020MS002195 (2020).
 71. Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E. & Knutti, R. Climate change now detectable from any single day of weather at global scale. *Nat. Clim. Chang.* **10**, 35–41 (2020).
 72. Ham, Y.-G. et al. Anthropogenic fingerprints in daily precipitation revealed by deep learning. *Nature* **622**, 301–307 (2023).
 73. de Vries, I. E., Sippel, S., Pendergrass, A. G. & Knutti, R. Robust global detection of forced changes in mean and extreme precipitation despite observational disagreement on the magnitude of change. *Earth Syst. Dyn.* **14**, 81–100 (2023).
 74. Watt-Meyer, O. et al. Ace: a fast, skillful learned global atmospheric model for climate prediction. 2310.02074. <https://arxiv.org/abs/2310.02074> (2023).
 75. Ghanem, R., Higdon, D. & Owhadi, H. *Handbook of Uncertainty Quantification* (Springer, 2017).
 76. Xu, L., Chen, N., Yang, C., Yu, H. & Chen, Z. Quantifying the uncertainty of precipitation forecasting using probabilistic deep learning. *Hydrol. Earth Syst. Sci.* **26**, 2923–2938 (2022).
 77. Bella, A., Ferri, C., Hernández-Orallo, J. & Ramírez-Quintana, M. in *Calibration of Machine Learning Models* 128–146 (IGI Global, 2010).
 78. Macherera, M. & Chimbari, M. J. A review of studies on community based early warning systems. *JAMBA* **8**, 206 (2016).
 79. Reichstein, M. et al. Early warning of complex climate risk with integrated artificial intelligence. *Nature Communications (accepted)* <https://doi.org/10.21203/rs.3.rs-4248340/v1> (2025).
 80. Corps, I. M. Libya flooding: Situation report #9 (2023).
 81. Tradowsky, J. S. et al. Attribution of the heavy rainfall events leading to severe flooding in western Europe during July 2021. *Clim. Chang.* **176**, 90 (2023).
 82. Attribution, W. W. Extreme downpours increasing in southern Spain as fossil fuel emissions heat the climate, accessed 5 November 2024. <https://www.worldweatherattribution.org/extreme-downpours-increasing-in-southern-spain-as-fossil-fuel-emissions-heat-the-climate/> (2024).
 83. Tojčić, I., Denamiel, C. & Vilibic, I. Performance of the Adriatic early warning system during the multi-meteorological event of 11–19 May 2020: an assessment using energy banners. *Nat. Hazard. Earth Syst. Sci.* **21**, 2427–2446 (2021).
 84. Yore, R. & Walker, J. Early warning systems and evacuation: rare and extreme vs frequent and small-scale tropical cyclones in the Philippines and Dominica. *Disasters* **45**, 691–716 (2020).
 85. Tamamadin, M. et al. Automation process to support an information system on extreme weather warning. *IOP Conference Series: Materials Science and Engineering* (IOP Publishing, 2020).
 86. AI, H. Ethics guidelines for trustworthy AI: high-level expert group on artificial intelligence. <https://wayback.archive-it.org/12090/20201227221227/https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (2018).
 87. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **1**, 389–399 (2019).
 88. Kochupillai, M., Kahl, M., Schmitt, M., Taubenböck, H. & Zhu, X. X. Earth observation and artificial intelligence: understanding emerging ethical issues and opportunities. *IEEE Geosci. Remote Sens. Mag.* **10**, 90–124 (2022).
 89. Ramos, M., van Andel, S. & Pappenberger, F. Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.* **17**, 2219–2232 (2013).
 90. Giuliani, M., Pianosi, F. & Castelletti, A. Making the most of data: an information selection and assessment framework to improve water systems operations. *Water Resour. Res.* **51**, 9073–9093 (2015).
 91. Camps-Valls, G., Tuia, D., Zhu, X. X. & Reichstein, M. *Deep learning for the Earth Sciences: A comprehensive Approach to Remote Sensing, Climate Science and Geosciences* (John Wiley & Sons, 2021).
 92. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019). **The authors discuss how deep learning can enhance process understanding in Earth system science, bridging data-driven approaches with traditional modeling.**

93. Iles, C. E. et al. The benefits of increasing resolution in global and regional climate simulations for European climate extremes. *Geosci. Model Dev.* **13**, 5583–5607 (2020).
94. Cheng, S. et al. Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA J. Autom. Sin.* **10**, 1361–1387 (2023).
95. Haynes, K., Lagerquist, R., McGraw, M., Musgrave, K. & Ebert-Uphoff, I. Creating and evaluating uncertainty estimates with neural networks for environmental-science applications. *Artif. Intell. Earth Syst.* **2**, 220061 (2023).
96. Li, W., Pan, B., Xia, J. & Duan, Q. Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts. *J. Hydrol.* **605**, 127301 (2022).
97. Farazmand, M. & Sapsis, T. P. Extreme events: mechanisms and prediction. *Appl. Mech. Rev.* **71**, 050801 (2019).
98. Zhang, M., Fernández-Torres, M. A. & Camps-Valls, G. Domain knowledge-driven variational recurrent networks for drought monitoring. *Remote Sens. Environ.* **311**, 114252 (2024).
99. Ronco, M. & Camps-Valls, G. Role of locality, fidelity and symmetry regularization in learning explainable representations. *Neurocomputing* **562**, 126884 (2023).
100. Roscher, R., Bohn, B., Duarte, M. F. & Garcke, J. Explain it to me: facing remote sensing challenges in the bio- and geosciences with explainable machine learning. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **5**, 817–824 (2020).
101. Beucler, T. et al. Climate-invariant machine learning. *Sci. Adv.* **10**, eadj7250 (2024).
102. West, H., Quinn, N. & Horswell, M. Remote sensing for drought monitoring & impact assessment: progress, past challenges and future opportunities. *Remote Sens. Environ.* **232**, 111291 (2019).
103. Zargar, A., Sadiq, R., Naser, B. & Khan, F. I. A review of drought indices. *Environ. Rev.* **19**, 333–349 (2011).
104. Dikshit, A. & Pradhan, B. Interpretable and explainable AI (XAI) model for spatial drought prediction. *Sci. Total Environ.* **801**, 149797 (2021).
105. Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D. & Salcedo-Sanz, S. Heat waves: physical understanding and scientific challenges. *Rev. Geophys.* **61**, e2022RG000780 (2023).
106. Teng, H., Leung, R., Branstator, G., Lu, J. & Ding, Q. Warming pattern over the northern hemisphere midlatitudes in boreal summer 1979–2020. *J. Clim.* **35**, 3479–3494 (2022).
107. Vautard, R. et al. Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends. *Nat. Commun.* **14**, 6803 (2023). **This study finds that heat extremes in Western Europe are intensifying more rapidly than climate models predict, attributed to changes in atmospheric circulation.**
108. Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P. & Bouchet, F. Deep learning-based extreme heatwave forecast. *Front. Clim.* **4**, 789641 (2022).
109. Fister, D., Pérez-Aracil, J., Peláez-Rodríguez, C., Del Ser, J. & Salcedo-Sanz, S. Accurate long-term air temperature prediction with machine learning models and data reduction techniques. *Appl. Soft Comput.* **136**, 110118 (2023).
110. van Straaten, C., Whan, K., Coumou, D., van den Hurk, B. & Schmeits, M. Correcting subseasonal forecast errors with an explainable ANN to understand misrepresented sources of predictability of European summer temperatures. *Artif. Intell. Earth Syst.* **2**, e220047 (2023).
111. Khan, M. I. & Maity, R. Hybrid deep learning approach for multi-step-ahead prediction for daily maximum temperature and heatwaves. *Theor. Appl. Climatol.* **149**, 945–963 (2022).
112. Happé, T. et al. Detecting spatio-temporal dynamics of western European heatwaves using deep learning. *Artif. Intell. Earth Syst.* <https://journals.ametsoc.org/view/journals/aies/aop/AIES-D-23-0107.1/AIES-D-23-0107.1.xml> (2024).
113. Gao, Z. et al. Earthformer: Exploring space-time transformers for Earth system forecasting. *Adv. Neural Inf. Process. Syst.* **35**, 25390–25403 (2022).
114. Sun, Y. Q. et al. Can AI weather models predict out-of-distribution gray swan tropical cyclones? *arXiv preprint arXiv:2410.14932* (2024).
115. Lopez-Gomez, I., McGovern, A., Agrawal, S. & Hickey, J. Global extreme heat forecasting using neural weather models. *Artif. Intell. Earth Syst.* **2**, e220035 (2023).
116. Pouplin, T., Jeffares, A., Seedat, N. & van der Schaar, M. Relaxed quantile regression: Prediction intervals for asymmetric noise. *arXiv preprint arXiv:2406.03258* (2024).
117. Trok, J. T., Davenport, F. V., Barnes, E. A. & Diffenbaugh, N. S. Using machine learning with partial dependence analysis to investigate coupling between soil moisture and near-surface temperature. *J. Geophys. Res. Atmos.* **128**, e2022JD038365 (2023).
118. Singh, J., Sippel, S. & Fischer, E. M. Circulation dampened heat extremes intensification over the midwest USA and amplified over western Europe. *Commun. Earth Environ.* **4**, 1–9 (2023).
119. Jones, M. W. et al. Global and regional trends and drivers of fire under climate change. *Rev. Geophys.* **60**, e2020RG000726 (2022).
120. El Garroussi, S., Di Giuseppe, F., Barnard, C. & Wetterhall, F. Europe faces up to tenfold increase in extreme fires in a warming climate. *npj Clim. Atmos. Sci.* **7**, 1–11 (2024).
121. Zhang, G., Wang, M. & Liu, K. Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. *Int. J. Disaster Risk Sci.* **10**, 386–403 (2019).
122. Li, F. et al. Attentionfire_v1.0: interpretable machine learning fire model for burned-area predictions over tropics. *Geosci. Model Dev.* **16**, 869–884 (2023).
123. Singh, H. et al. Trending and emerging prospects of physics-based and ml-based wildfire spread models: a comprehensive review. *J. For. Res.* **35**, 1–33 (2024).
124. Fromm, M., Servranckx, R., Stocks, B. J. & Peterson, D. A. Understanding the critical elements of the pyrocumulonimbus storm sparked by high-intensity wildland fire. *Commun. Earth Environ.* **3**, 1–7 (2022).
125. Salas-Porras, E. D. et al. *Identifying the Causes of Pyrocumulonimbus (PyroCb)* <http://arxiv.org/abs/2211.08883> (2022).
126. Jonkman, S. Global perspectives on loss of human life caused by floods. *Nat. Hazards* **34**, 151–175 (2005).
127. Cornwall, W. Europe's deadly floods leave scientists stunned. *Science* **373**, 372–373 (2021).
128. Mohr, S. et al. A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe—Part 1: Event description and analysis. *Nat. Hazards Earth Syst. Sci.* **23**, 525–551 (2023).
129. Xie, K. et al. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *J. Hydrol.* **603**, 127043 (2021).
130. Kraft, B., Jung, M., Körner, M., Koirala, S. & Reichstein, M. Towards hybrid modeling of the global hydrological cycle. *Hydrol. Earth Syst. Sci. Discuss.* **2021**, 1–40 (2021).
131. Kochkov, D. et al. Machine learning-accelerated computational fluid dynamics. *Proc. Natl Acad. Sci. USA* **118**, e2101784118 (2021).
132. Lutjens, B. et al. Physically-consistent generative adversarial networks for coastal flood visualization. <https://doi.org/10.48550/arXiv.2104.04785> (2023).

Acknowledgements

G.C.-V., M.-A.F.-T., K.-H.C., A.P., and M.R. acknowledge the European Research Council (ERC) support under the ERC Synergy Grant USMILE (grant agreement 855187). G.C.-V., M.-A.F.-T., M.R., T.H., M.G., T.W., and S.S. thank the H2020 XAIDA project (grant agreement 101003469). G.C.-V., M.G., and O.-J.P.-V. thank the Horizon project ELIAS (grant agreement 101120237). G.C.-V., M.R., and S.S. acknowledge the Horizon project

AI4PEX (grant agreement 101137682). A.C., M.G., S.S.-S., and J.P.A. thanks the H2020 CLINT project (grant agreement 101003876). S.S.-S. and J.P.-A. acknowledge the support from project PID2023-150663NB-C21 of the Spanish Ministry of Science, Innovation and Universities (MICINNU). I.P., I.P., and S.K. acknowledge the Horizon Europe project MeDiTwin (grant agreement 101159723) and the H2020 project Deep-Cube (grant agreement 101004188). S.S. thanks the climXtreme project (Phase 2, project PATTETA, grant number 01LP2323C), funded by the German Federal Ministry of Education and Research. K.W. acknowledges the support from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Gottfried Wilhelm Leibniz Prize awarded to Veronika Eyring (Reference number EY 22/2-1).

Author contributions

This piece originated from discussions in an ELLIS.eu workshop titled “Extreme Event Detection, Analysis, and Explanation in Earth and Climate Sciences” held in València, Spain, in October 2023. ELLIS is the biggest AI network of excellence in Europe. The workshop gathered more than 40 experts in the intersection of AI, climate sciences, and extreme events for a week. G.C.-V. designed and structured the study with the help of M.-A.F.-T. and K.-H.C. G.C.-V. wrote the first version of the manuscript with the help of all coauthors. G.C.-V. supervised the study. M.-A.F.-T., A.H., A.C., A.P., O.-J.P.-V., and M.G. worked on detection and forecasting. C.R., F.M., I.Pr., K.W., and I.Pa. worked on impact assessment. M.-A.F.-T., K.-H.C., M.G.-C., A.H., S.S.-S., M.R., O.-I.P., O.-J.P.-V., and M.Ra. worked on extreme event understanding (XAI, UQ). G.C.-V., A.C., and M.G. worked on the operationalization of the chain. M.Re., G.C.-V., A.C., S.S. worked on communication of risk, ethics, and decision-making. G.C.-V., S.O., and A.C. worked on the data and model integration challenges. M.-A.F.-T., C.R., M.G.-C., F.M., M.G., and T.W. worked on the drought case study. T.H., M.D.M., S.S.-S., J.P.-A. worked on the heat-waves case study. I.Pr., S.K. worked on the wildfires case study. G.C.-V. and M.R. worked on the floods case study. 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/s41467-025-56573-8>.

Correspondence and requests for materials should be addressed to Gustau Camps-Valls.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://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

¹Image Processing Laboratory, Universitat de València, València, Spain. ²Chair of Data Science in Earth Observation, Technical University of Munich, Munich, Germany. ³Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milano, Italy. ⁴RFF-CMCC European Institute on Economics and the Environment, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Milan, Italy. ⁵Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany. ⁶University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany. ⁷Max Planck Institute of Biogeochemistry, Jena, Germany. ⁸ELLIS, ELLIS Unit Jena, Jena, Germany. ⁹Institute for Earth System Science & Remote Sensing, Leipzig University, Leipzig, Germany. ¹⁰Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig, Germany. ¹¹National Technical University of Athens, Athens, Greece. ¹²National Observatory of Athens, Athens, Greece. ¹³Archimedes/Athena Research Center, Athens, Greece. ¹⁴Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain. ¹⁵German Aerospace Center (DLR), Institute for Data Science, Jena, Germany. ¹⁶Department of Mathematical and Electrical Engineering, IMT Atlantique, Lab-STICC, UMR CNRS 6285 & INRIA team odyssey, Brest, France. ¹⁷Institute for Meteorology, Leipzig University, Leipzig, Germany. ¹⁸Institute for Environmental Studies, VU Amsterdam, Amsterdam, The Netherlands.

✉ e-mail: gustau.camps@uv.es