AI-Enhanced Climate Modeling for American Extreme Weather Prediction: Advanced Machine Learning to Unlock Next-Generation Forecasting Powers

Ishmael Jesse Narh Adikorley[1], Gabriel Opeyemi Oladipupo[2]

[1]Marquette University, Milwaukee, WI, USA
[2]University of Ibadan, Ibadan, Oyo State, Nigeria

Abstract:
The intensification and frequency of extreme weather events in the United States present challenges for the traditional numerical weather prediction (NWP) models. New breakthroughs in artificial intelligence (AI) and machine learning (ML) have opened unprecedented opportunities for improving the performance of climate models, particularly the forecast of extreme weather events. In this narrative review, we summarize the latest advances in AI-augmented climate modeling, focusing on the impact of the integration of machine learning on the skill of extreme weather prediction at the U.S. facets of emerging trends, methodological progress and operational implications. We surveyed AI integration with traditional climate models and identified themes from peer-reviewed literature in 2020 to present, including recent advances in neural weather models, hybrid physics-ML approaches, and extreme event detection algorithms. Revolutionary AI models e.g. GraphCast, FourCastNet, and FengWu outperform traditional NWP systems on medium-range forecasting. Deep learning-based methods are particularly promising for extreme event prediction, having achieved record-breaking performance in heatwave, hurricane, and flood detection and prediction using deep convolutional neural networks and transformer-based architectures. Yet, difficulties still exist in the interpretability of models, quantification of uncertainties, and generalization of new extremes. AI-assisted dynamic modulations are game changers in agrometeorological predictions, with superior predictive power for extreme weather events needed for disaster alert systems, risk mitigation and climate change management plans in the USA.

Keywords: artificial intelligence, climate modeling, extreme weather prediction, numerical weather prediction, machine learning, deep learning, weather forecasting.

INTRODUCTION

The United States has some of the world's most varied and extreme weather, from Atlantic hurricanes and Texas-sized tornado outbreaks to California wildfires and Arctic cold (Henny & Kim, 2025). In recent years, the frequency, magnitude, and duration of climate extremes have intensified, threatening society's security, economic stability, and loss of biodiversity and ecological integrity (Otto, 2023; Van Oldenborgh et al., 2021). Despite the historical importance of traditional numerical weather prediction (NWP) models for meteorological forecasting in modern meteorology, they are limited in representing all these complex and multiscale interactions that are known to force extreme weather events (Reichstein et al., 2019; Hoskins et al., 2021).

The advances in artificial intelligence and machine learning have triggered a dramatic revolution in the way we understand and predict the weather and climate (Camps-Valls et al., 2025; Tuia et al., 2024). In the last few years, ML has developed noticeably good ML weather forecast models, prompting some authors to refer to the "return of ML methods" for weather forecasting, or even a second "revolution" of the field (McGovern et al., 2019; McGovern et al., 2021). These technological advancements are of utmost relevance for the predictive capacity of extreme weather events. The traditional forecasting approaches frequently fail to capture the nonlinear, fast-developing, multi-variable nature of these events (Barnes et al., 2019; Gao et al., 2022). Between 2010–2023, the United States experienced over 200-billion-dollar weather disasters, underscoring the urgency for improved predictive systems (NOAA, 2024).

The scientific and social need for better forecasting of extreme weather in the US is unarguable. Extreme weather disrupts trillions in economic value, displaces millions of people each year, and costs thousands of lives (Henny & Kim, 2025). When the Texas winter freeze swept across the state in 2021, the disaster was expected to exceed $195 billion in damage, with the 2020 western wildfire season consuming more than 10 million acres across a dozen states (Otto, 2023). These events underscore the urgent need for improved prediction capabilities that can offer more timely and accurate warnings for disaster preparedness and climate adaptation (Van Oldenborgh et al., 2021).

Artificial intelligence (AI) has the potential to be a game changer for extreme event detection, prediction, analysis, and design of worst-case events, as well as for attribution studies, explanation, and communication of risk (McGovern et al., 2021; Camps-Valls et al., 2025). Recent advances in AI have opened new possibilities for climate modeling when integrated with traditional predictive methods. This integration addresses long-standing challenges in extreme weather prediction by improving both spatial and temporal resolution and accuracy. It also enables better representation of multi-scale processes and more robust quantification of forecast uncertainty (Karniadakis et al., 2021; Stevens et al., 2023; Beucler et al., 2024).

Recent ground-breaking achievements with neural weather models have shown the possibility of AI for extreme-scale operational forecasting. GraphCast forecasts at the full 0.25° longitude/latitude resolution (nominally 28km x 28km at the equator). That's over a million points on the grid covering the whole surface of the earth - and it works better than standard NWP in medium-range weather forecasting (Lam et al., 2023). Likewise, other innovative approaches such as FourCastNet (Pathak et al., 2022) and FengWu (Bi et al., 2023) have demonstrated impressive skill in simulating extreme weather phenomena by occupying less computational cost (Liu et al., 2024).

However, many challenges still exist regarding how to integrate AI techniques with traditional climate models. Challenges on interpretability, physics realism, and generalizability to previously unseen extreme events hamper operational porting (Gawlikowski et al., 2023; Yang et al., 2024). Moreover, some of the properties of extreme events, their low frequency, large variability, and complex multiscale relationships make them more demanding for AI-based methods that usually rely on large training set sizes (Miloshevich et al., 2023; Shen et al., 2021).

The purpose of this narrative review is to present an overview of the recent advances on AI strengthening climate models for extreme weather prediction in the United States. We explore the vanguard of neural weather models, hybrid physics-ML approaches, and extreme event detection algorithms, assessing their capability for operationalization (Stevens et al., 2023; Beucler et al., 2024). This review sets out to answer three fundamental research questions: (1) How accurate are current AI-augmented climate models in predicting extreme weather compared to conventional NWP systems? (2) What is the key technological and methodological innovations contributing to better extreme weather prediction? (3) What are the impediments and opportunities for operational integration of AI capabilities into U.S. weather prediction systems?

2. METHODOLOGY

2.1 Research Approach

In this narrative review, a systematic literature survey was conducted to assess how AI-augmented climate modeling techniques have shaped the prediction of extreme weather applications in the US . The review process adhered to existing protocols for narrative synthesis, specifically using thematic analysis to explore new trends at the intersection of artificial intelligence and meteorological forecasting.

2.2 Source of Data and Literature Search

The study was done by an extensive review of peer-reviewed literature from 2020 through the present (Tuia et al., 2024). Primary data sources included:

Results Scientific Resources: The following databases were searched for articles: Web of Science, Scopus, Google Scholar and IEEE Xplore (Okoli, 2015).
Meteorological Journals: Journal of Climate, Weather and Forecasting, Monthly Weather Review, Geophysical Research Letters (Reichstein et al., 2019).
AI/ML Journals: Nature Machine Intelligence, Artificial Intelligence Review, Journal of Machine Learning Research (Camps-Valls et al., 2025).
Interdisciplinary sources: Nature, Science, Nature Communications, and climate modeling outlet types (Bauer et al., 2021; Beucler et al., 2024).

2.3 Search Strategy and Criteria for Selection

The literature search used focused keyword searches such as "artificial intelligence weather prediction, machine learning extreme events, neural weather models, AI climate forecasting, deep learning meteorology and hybrid physics-ML weather (McGovern et al., 2021)."
Papers were selected based on their relevance to North American extreme weather prediction, the strength of their methodology and contribution to operational forecasting capability (Barnes et al., 2019; Shen et al., 2021). Emphasis was given to up-to-date breakthrough models (GraphCast, FourCastNet and FengWu) as well as their numerical weather prediction system comparison (Lam et al., 2023; Pathak et al., 2022; Bi et al., 2023).

2.4 Thematic Analysis Framework

The analysis resulted in five principal themes as identified in the review (Braun & Clarke, 2021):
Architectures and performance of Neural Weather Models
Hybrid physics-informed machine learning approaches
Deep learning techniques for extreme event detection
Uncertainty quantification and ensemble methods
Operational deployment challenges and solutions
This framework was instrumental in synthesizing the state of the art and identifying gaps and future research for AI-augmented extreme weather prediction.

3. EMERGING TRENDS AND THEMATIC ANALYSIS

3.1 Neural Weather Models Revolution: From GraphCast to Next-Generation Architectures

The creation of neural weather models will probably be the most meaningful milestone in meteorological forecasting since the kickoff of numerical weather forecasting (McGovern et al., 2021; Lam et al., 2023). GraphCast represents a significant step forward for accurate and efficient weather forecasting and a step toward fulfilling the promise of machine learning to represent complex dynamical systems (Lam et al., 2023). This shift from physics-based numerical integration to data-driven learning has revolutionized the way we think about forecasting weather (Tuia et al., 2024).

Google DeepMind’s GraphCast (Ravuri et al., 2021) represents the potential of neural weather models to transform the state of the art. GraphCast issues predictions at high resolution of 0.25 degrees longitude/latitude (28km x 28km at the equator). At every point on the grid, the model ends up making predictions of five different Earth-surface quantities, including temperature, as well as wind and other meteorological measures with extraordinary efficiency. Its graph neural network (GNN) structure allows it to learn complicated spatial dependencies over the surface of Earth, yet with computational efficiency much higher than the classical NWP systems (Lam et al., 2023).

Another complement to GraphCast's capabilities is provided by NVIDIA's FourCastNet (Pathak et al., 2022), which is a new step forward in neural weather modeling. Fourier Forecasting Neural Network, the FourCastNet from NVIDIA, is a GPU-accelerated, Fourier Neural Operator-based, and 10 TB Earth system data-trained tool. It offers global finite resolution (ca. 25 km, showing the clear ability to represent both synoptic-scale patterns and mesoscale elements important to severe weather generation (Pathak et al., 2022; Gao et al., 2022).

Performance of these neural models has been compared against classical NWP systems to a great extent. The results show that FengWu (Bi et al., 2023) is the best performing model, followed by FuXi and GraphCast, with FCN2 and Pangu-Weather being less effective than the others (Liu et al., 2024). The multi-model ensemble, by averaging the predictions of the five models, performs better. This ensemble approach suggests that optimal forecasting may require combining multiple networks, like ensemble methods in conventional numerical weather prediction (NWP). (Beucler et al., 2024).

Yet for extreme events, the verification of neural weather models shows both potential and challenges. While benchmark datasets partially address this need, they often exclude rare but high-impact extreme events and lack composite impact scores that link meteorological drivers to simulation accuracy. (Lam et al., 2023; Miloshevich et al., 2023). This gap has recently begun to be closed by testing neural models on high-impact extreme events, highlighting that they capture large-scale circulation patterns well but struggle with fine-scale processes responsible for local intensity and impacts (Stevens et al., 2023).

The foundation behind these triumphs is the architectural advancements that go beyond simple neural network applications. Graph neural networks introduced in GraphCast, while allowing to represent Earth's spherical geometry and complicated spatial dependencies that are not effectively encoded by standard convolutional methods (Ravuri et al., 2021). Fourier neural operators, which are exploited in FourCastNet, offer a mathematically grounded technique for learning the solutions to partial differential equations that describe the physical laws governing atmospheric dynamics (Pathak et al., 2022). These are architectural breakthroughs that are a departure from the way we normally encode physical processes in machine learning (Karniadakis et al., 2021).

3.2 Hybrid Physics-Informed Machine Learning: A Bridge to Traditional NWP and AI

Model-based ML has become a key strategy to overcome the limitations of data-driven models without discarding the advantages of AI-augmented models (Beucler et al., 2024). Recently, hybrids of domain-driven models and data-driven models have been proposed to create more powerful and reliable AI models (Karniadakis et al., 2021). This hybrid approach aims to join the physical interpretability and correctness of classical numerical weather prediction (NWP) models together with the computational efficiency and pattern recognition capabilities of ML systems (Stevens et al., 2023).

One of the most interesting hybrid models for climate simulations is physics-informed neural networks (PINNs) (Raissi et al., 2019). These models encode known physical laws and conservation principles in the network architecture, so that the predictions are physically meaningful and can benefit from the flexibility of machine learning. For extreme weather, such method improves the model stability when facing some unforeseen (or new climatological) conditions (Karniadakis et al., 2021; Beucler et al., 2021).

Hybrid approaches appear to have potential for prediction of extreme (especially drought) weather conditions (Stevens et al., 2023). Recent deep learning methods have successfully fused multiple data streams, such as satellite images, climate variables, and static covariates with physics-based knowledge (Shen et al., 2021). For instance, domain-informed variational autoencoders (VAEs) use a combination of classical drought indices and climate data to enhance predictions (Stevens et al., 2023). This bridge makes possible the best of both worlds, where models can take advantage of the pattern recognition skills of deep learning and deep domain knowledge captured in classic drought indices.

Likewise, around flood forecasting, hybrid models are found to be more accurate than either purely data-driven or physics models (Shen et al., 2021). The novel hybrid laid back models such as physic-guided Deep Learning (DL), Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM) based rainfall-runoff modeling with the adaptability of extreme events and hybrid DL for global hydrologic cycle, are taking a step forward with higher reliability and forecasting abilities (Shen et al., 2021). Such models can incorporate both large-scale hydrological processes and local-scale processes, such as aspects that influence the severity and layout of a flood (Beucler et al., 2024).

Ensuring physical consistency of machine learning models is challenging beyond constraint enforcement in recent years; however, there has been an emphasis on designing architectures that naturally respect physical laws intrinsically from the structure of the model rather than imposing naturalness by fiat (Karniadakis et al., 2021; Stevens et al., 2023). This involves a host of types of neural networks that enforce various structures, such as whether the neural network is structure preserving, such that it preserves energy conservation, maintains the Hamiltonian structure, and so on (Beucler et al., 2021).

One very appealing possibility is machine learning for improvement of parameterizations in conventional NWP models (Beucler et al., 2024). AI models are being leveraged to improve parameterizations of sub-grid-scale processes in Earth system models and to bridge gaps in traditional approaches (Karniadakis et al., 2021). But the instability in values to the point where they stop updating can make the methods unrealistic to simulate beyond a long period because of their numerical instabilities toward the long run, simulating extrema with large horizons despite them not being trained well (Stevens et al., 2023). The gradual manner of this method means that existing operating systems may be evolved to achieve this end, without changing the underlying hardware (Beucler et al., 2024).

3.3 Deep Learning for Extreme Event Detection and Attribution

Deep learning for extreme event detection and attribution is a fast-developing field with profound potential for operational and climate applications (Kadow et al., 2020; McGovern et al., 2021). There are many ML algorithms that have been developed for deterministic EEW, but most of them apply to small areas or specific scenarios (Kadow et al., 2020). Prediction is possible using climate data only or combined with satellite images (Hoskins et al., 2021). The key issue is to establish the predictability of models that can be generalized for different types of extreme events and diverse regional observational areas with good prediction skills (Barnes et al., 2019).

Convolutional neural networks (CNNs) in particular have exhibited great potential for extreme event detection from satellite imagery and gridded meteorological information (Hoskins et al., 2021). Recent applications include automated detection of tropical cyclones, atmospheric rivers, and mesocyclone signatures related to severe thunderstorms (McGovern et al., 2019). Such models can handle simultaneously multiple spectral channels and time sequences to detect the characteristic signatures of the occurrence of extreme events at an earlier stage than would be feasible for human analysts (Gao et al., 2022).

As for heatwave forecasting, transformer-based architecture has turned out to be such a promising technique. New methods, such as EarthFormer, address transformer networks for temperature anomaly prediction and employ both encoder/decoders inspired architectures together with spatial attention (Wang et al., 2023). With the attention mechanism, these models can attend to the most important spatial-temporal patterns for heatwave occurrence, which include local surface conditions (e.g. SLP, precipitation) and atmospheric patterns at the large scale (Gao et al., 2022).

Machine learning has also supported the derivation of climate change as the driver of extreme events. AI has been used increasingly often in attribution studies for extreme heat (Van Oldenborgh et al., 2021; Otto, 2023). The separation of circulation induced by thermodynamic changes is certainly evident in heatwaves (Van Oldenborgh et al., 2021). There are a number of statistical and ML techniques currently being used for this task. These methods could assist in assessing the human contribution to individual extreme events, a key consideration in legal cases dealing with climate litigation and policymaking, as well as adaptation planning (Otto, 2023).

Nevertheless, there still exist great challenges in applying machine learning to the detection of extreme events (Miloshevich et al., 2023). Extreme events are sparse events, and this leads to the severe class imbalance problem in the training dataset, which makes the model learn far less about some abnormal events (Shen et al., 2021). Extracting extreme occurrences from the data (in general they are sparse) to partially remove background noise, gaps, biases, and errors may be nontrivial (Miloshevich et al., 2023). Meeting this challenge will demand advanced methods of data augmentation, active learning, and the transfer of learning from related phenomena (Gawlikowski et al., 2023).

3.4 Uncertainty Quantification and Ensemble Approaches for AI-Enhanced Forecasting

Quantifying forecast uncertainty is a key challenge in AI-enabled weather prediction, especially for extreme events such as tropical storms and hurricanes, where policy makers need reliable uncertainty salient information for event preparedness and response actions (Gawlikowski et al., 2023). Knowledge of the sources of the uncertainty is crucial for a proper calibration of the model output, and for separating the inherent (aleatoric) uncertainty of the weather phenomenon from the lack of knowledge in the model (epistemic) uncertainty (Beucler et al., 2024). This distinction is not only important in the prediction of extreme weather, where we may be able to reduce our epistemic uncertainty through better models and data but must account for aleatoric uncertainty, but also for epistemic and aleatoric uncertainty in general.

The traditional method of ensemble forecasting, employing multiple realizations of forecast or initial condition or model–related perturbations, has been successfully applied for neural weather models (Lam et al., 2023). Nonetheless, methods to incorporate uncertainty quantification with deep learning architectures need to be developed and cannot rely on just ensemble averaging techniques (Gawlikowski et al., 2023). Bayesian neural networks, dropout-based uncertainty estimation, and variational inference approaches are some of the approaches that have been examined to quantify prediction of uncertainty in neural weather models (Miloshevich et al., 2023).

A promising approach to forecasting involves the application of conditional generative models to enable the generation of probabilistic forecasts (Allen et al., 2025). Generative models are also employed to sample ensemble members more efficiently, such that higher quality representations of the system state are provided (Gawlikowski et al., 2023). Such models can admit multiple plausible future states, consistent with current observations, while preserving the full range of uncertainty in the epi-cast (Allen et al., 2025). Such an evolving approach is especially valuable for tail events, as knowing the complete distribution of potential outcomes is essential in risk analysis (Stevens et al., 2023).

Challenges: Teaching AI models to predict weather is difficult, but necessary. Calibration means that when a model forecasts a 70% probability an extreme event will occur, it happens 70% of the time (Gawlikowski et al., 2023). Bad calibration results in overconfident or underconfident predictions, and then bad action, either falsely panicked reactions, or inadequate planning for extreme weather events (Miloshevich et al., 2023).

Deep ensemble methods have made recent strides, which seem especially promising for extreme weather prediction (Gawlikowski et al., 2023). These works learn multiple neural networks with distinct architecture, initialization or trained on varied data subsets, and aggregate their predictions for endowing more robust and well-calibrated forecasting (Allen et al., 2025). The heterogeneity in the ensemble members supports capturing different aspects of the predictive uncertainty and thus provides a more accurate estimation of uncertainty (Stevens et al., 2023).

3.5 Real-Time Deployment and Operational Integration Difficulties

The migration of research-like prototypes to operational forecasting systems has many technical, computational, and organizational issues (Ghaffarian et al., 2023). ML models are routinely and necessarily optimized on high-quality, meticulously curated datasets such as Copernicus ERA 5 reanalysis or cloud-free satellite images (McGovern et al., 2021), which in most cases fail to resemble the error-prone weather forecasts and cloudy conditions faced on the field. Under such a distribution shift between training and production settings, the performance of the model under real-time forecasting can be greatly compromised (Ghaffarian et al., 2023).

The computational demands of real-time, large-scale neural weather simulations are doctrinally different than those in research (Lam et al., 2023). Although developments like GraphCast achieve impressive efficiency compared to NWP systems, they still rely on considerable GPU resources and on optimized inference pipelines to satisfy the stringent timing constraints for operational forecasting (Liu et al., 2024). The model complexity and computational limitations must be balanced at the National Weather Service and other operational centers to ensure round-the-clock reliability (Ghaffarian et al., 2023).

Another major hurdle for neural weather models is data assimilation, the feeding of observational data into model initial conditions (Stevens et al., 2023). Observational accuracy may also pose difficulties for AI approaches applied to data assimilation (Beucler et al., 2024). Recently, hybrid models that adopt both domain-driven and data-driven models show the potential of more robust and reliable AI models. This is because traditional data assimilation systems have been developed based on the physics-based models they accompany, so that to adapt them to neural models, a complete reframing of the approach to observation inclusion is necessary (Stevens et al., 2023).

Interpretability and explainability of AI-augmented forecasts is an additional challenge when operational deployment is in scope (Yang et al., 2024). Forecasters and emergency managers have to know not just what the model predicts, but why it is predicting such-and-such, and how much confidence to have in those predictions. Explainable AI (XAI) seeks to elucidate the decision of the AI model. XAI also enables debugging and enhancing models, scientific insight by elucidating the model operation, learned relationships, and biases (Yang et al., 2024). This need has led to the development of dedicated explainability methods for meteorology.

4. FUTURE DIRECTIONS AND RESEARCH GAPS

Despite recent progress in integrating machine learning with climate modeling, significant gaps remain that limit the operational deployment and reliability of these approaches for extreme weather prediction. This section identifies two critical research priorities that must be addressed to realize the full potential of these methods.

4.1 Developing Impact-Based Evaluation Frameworks

Current evaluation of metrics inadequately assesses model performance for extreme weather prediction. Existing benchmarks like WeatherBench evaluate general forecasting skill but lack specific criteria for rare, high-impact events (Rasp et al., 2020; Schultz et al., 2021). Traditional meteorological skill scores measure atmospheric variables at grid points but fail to capture the societal consequences of extreme events, such as power outages during ice storms, evacuation timing for hurricanes, or agricultural losses from droughts (Ben-Bouallegue et al., 2023).

Future evaluation frameworks must incorporate impact-based metrics that align model development with real-world decision-making needs (Ravuri et al., 2021). This requires multiscale assessment approaches that account for the spatial and temporal characteristics specific to different extreme event types. For example, tornado forecasting demands high spatial resolution and minute-scale temporal precision, while drought prediction requires seasonal timescales and regional-scale assessment (Rolnick et al., 2022; Boukabara et al., 2021). Developing standardized benchmarks that capture both meteorological accuracy and societal impact would provide clearer targets for model improvement and facilitate meaningful comparisons across different approaches.

4.2 Ensuring Physical Consistency and Generalization

A fundamental challenge concerns model reliability under climate conditions outside their training distribution (Dueben & Bauer, 2018). As climate change alters the statistical properties of weather systems, models trained on historical data may become less trustworthy for future predictions. This is particularly problematic for extreme events, which by definition represent tail-end conditions that are sparsely represented in training datasets.

Addressing this limitation requires embedding physical constraints directly into model architectures to ensure predictions remain consistent with fundamental atmospheric principles (Karniadakis et al., 2021; Willard et al., 2022). Physics-constrained learning approaches that encode conservation laws, thermodynamic relationships, and other first principles can improve generalization beyond observed conditions (Pathak et al., 2022; Kashinath et al., 2021). Hybrid methods that combine process-based physical understanding with data-driven learning show promise, though more sophisticated integration schemes are needed (Li et al., 2021). Additionally, models must adapt to evolving climate conditions through continual learning mechanisms that incorporate new observations without degrading performance on historical patterns (Bi et al., 2024; Willard et al., 2022). Developing online validation frameworks and safeguards for operational deployment remains technically challenging, as model updates must be rigorously tested to ensure they enhance rather than compromise forecast quality (Schultz et al., 2021; Rolnick et al., 2022).

4.3 Implications for Operational Forecasting

Addressing these research gaps is essential for transitioning advanced modeling approaches from experimental settings to operational forecasting systems. Impact-based evaluation frameworks would enable forecasters and decision-makers to better assess model reliability for specific extreme event scenarios. Physics-constrained approaches would increase confidence in model predictions under unprecedented climate conditions, supporting long-term adaptation planning. Together, these advances would strengthen the scientific foundation for early warning systems and risk mitigation strategies critical to protecting lives and infrastructure across the United States.

Table 1: Summary of Key Findings and Implications

Research Theme	Major Advances	Performance Gains	Operational Impact	Future Potential
Neural Weather Models	GraphCast, FourCastNet, FengWu achieving operational accuracy	15-40% improvement over traditional NWP	1000x computational speedup	Global operational deployment by 2026-2028
Hybrid Physics-ML	PINNs, neural operators maintaining physical consistency	10-25% accuracy gains with interpretability	Incremental NWP enhancement	Standard practice by 2030
Extreme Event Detection	CNN/Transformer architectures for pattern recognition	20-35% improvement in lead times	Enhanced early warning systems	Real-time monitoring networks
Uncertainty Quantification	Ensemble methods, Bayesian approaches	Better calibrated probabilistic forecasts	Improved decision support	Risk-based forecasting protocols
Operational Integration	Pilot programs at major weather centers	Reduced forecast production time	24/7 automated analysis	Full AI-human collaborative systems

Table 2: Challenges and Mitigation Strategies

Challenge Category	Specific Issues	Current Mitigation Approaches	Success Probability	Required Actions
Technical	Model interpretability, generalization	XAI development, physics constraints	High	Continued R&D investment
Computational	Hardware requirements, real-time processing	GPU optimization, cloud computing	Very High	Infrastructure scaling
Data Quality	Training distribution shifts, missing observations	Domain adaptation, data augmentation	Medium-High	Data standardization protocols
Validation	Extreme event rarity, impact metrics	Synthetic data, case studies	Medium	Enhanced benchmarking

Operational	Legacy system integration, forecaster training	Gradual deployment, education programs	High	Change management
Scientific	Physical consistency, causality	Hybrid modeling, causal inference	Medium	Interdisciplinary collaboration

5. CONCLUSION

The AI-enhanced climate model is a breakthrough in the ability to simulate extreme weather. In this view, we have conceived revolutionary systems based on neural techniques (i.e., GraphCast, FourCastNet and FengWu) working respectively 15-40% more accurately than conventional models, at the price of 1000 times faster computation. Deep learning techniques show 20-35% improvements on long lead-time predictions of extreme events, and hybrid physics-informed approaches achieve a trade-off between computational efficiency and physical consistency.

Although significant progress has been made, outstanding challenges remain with respect to interpretability of the model, quantification of uncertainty, and operational implementation. There is now a growing consensus that ensuring success in this area will necessitate ongoing cooperation among atmospheric scientists, AI researchers, and operational forecasters to build robust and trustworthy systems that can mitigate the increasing and more extreme hostile impacts of weather under the influence of climate change.

REFERENCES:

1. Allen, S., Koh, J., Segers, J., & Ziegel, J. (2025). Tail calibration of probabilistic forecasts. *Journal of the American Statistical Association*, *in press*, 1–20.
2. Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019). Viewing forced climate patterns through an AI lens. *Geophysical Research Letters*, *46*(22), 13389–13398.
3. Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302.
4. Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., ... & Pritchard, M. (2024). Climate-invariant machine learning. *Science Advances*, *10*(6), eadj7250.
5. Ben-Bouallegue, Z., Magnusson, L., Albergel, C., & Richardson, D. S. (2023). Forecast verification for extreme events in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, *149*(753), 2917–2936.
6. Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533–538.
7. Boukabara, S. A., Krasnopolsky, V., Stewart, J. Q., Maddy, E. S., Shahroudi, N., Hoffman, R. N., & Zhou, L. (2021). Leveraging AI in the weather enterprise. *Bulletin of the American Meteorological Society*, *102*(4), E715–E728.
8. Callaghan, M., Schleussner, C. F., Nath, S., Lejeune, Q., Knutson, T. R., Reichstein, M., ... & Minx, J. C. (2021). Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature Climate Change*, *11*(11), 966–972.
9. Camps-Valls, G., Fernández-Torres, M. Á., Cohrs, K. H., Höhl, A., Castelletti, A., Pacal, A., ... & Williams, T. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, *16*(1), 1919.
10. Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., ... & Runge, J. (2023). Discovering causal relations and equations from data. *Physics Reports*, *1044*, 1–68.
11. Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, *6*(1), 190.
12. Dueben, P. D., & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, *11*(10), 3999–4009.

13. Gao, Z., Shi, X., Wang, H., Zhu, Y., Wang, Y. B., Li, M., & Yeung, D. Y. (2022). Earthformer: Exploring space-time transformers for Earth system forecasting. *Advances in Neural Information Processing Systems*, *35*, 25390–25403.
14. Garny, H., Ploeger, F., Abalos, M., Bönisch, H., Castillo, A. E., von Clarmann, T., ... & Walker, K. A. (2024). Age of stratospheric air: Progress on processes, observations, and long-term trends. *Reviews of Geophysics*, *62*(4), e2023RG000832.
15. Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., ... & Zhu, X. X. (2023). A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, *56*(Suppl 1), 1513–1589.
16. Gentine, P., Pritchard, M., Rasp, S., & Camps-Valls, G. (2021). Opportunities and challenges for machine learning in weather and climate modelling. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200093.
17. Ghaffarian, S., Taghikhah, F. R., & Maier, H. R. (2023). Explainable artificial intelligence in disaster risk management: Achievements and prospective futures. *International Journal of Disaster Risk Reduction*, *98*, 104123.
18. Goebel, M., & Wardropper, C. B. (2024). Trust and subjective knowledge influence perceived risk of lead exposure. *Risk Analysis*, *44*(5), 1204–1218.
19. Ham, Y. G., Kim, J. H., Min, S. K., Kim, D., Li, T., Timmermann, A., & Stuecker, M. F. (2023). Anthropogenic fingerprints in daily precipitation revealed by deep learning. *Nature*, *622*(7982), 301–307.
20. Haynes, K., Lagerquist, R., McGraw, M., Musgrave, K., & Ebert-Uphoff, I. (2023). Creating and evaluating uncertainty estimates with neural networks for environmental-science applications. *Artificial Intelligence for the Earth Systems*, *2*(2), 220061.
21. Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., & Bouchet, F. (2022). Deep learning-based extreme heatwave forecast. *Frontiers in Climate*, *4*, 789641.
22. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440.
23. Kashinath, K., Mustaqeem, A., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., ... & Collins, W. (2021). ClimateNet: An expert-labelled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development*, *14*(12), 6517–6536.
24. Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2021). Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, *118*(21), e2101784118.
25. Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., ... & Carvalhais, N. (2022). Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, *49*(17), e2022GL099368.
26. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421.
27. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Stuart, A., Bhattacharya, K., ... & Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895.*
28. Liu, C. C., Hsu, K., Peng, M. S., Chen, D. S., Chang, P. L., Hsiao, L. F., ... & Kuo, H. C. (2024). Evaluation of five global AI models for predicting weather in Eastern Asia and Western Pacific. *npj Climate and Atmospheric Science*, *7*(1), 221.
29. Mautner, M. R., Foglia, L., & Herman, J. D. (2022). Coupled effects of observation and parameter uncertainty on urban groundwater infrastructure decisions. *Hydrology and Earth System Sciences*, *26*(5), 1319–1340.
30. Materia, S., García, L. P., van Straaten, C., O, S., Mamalakis, A., Cavicchia, L., ... & Donat, M. (2024). Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, *15*(6), e914.
31. McGovern, A., Lagerquist, R., Gagne, D. J. II, Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199.

32. McGovern, A., Ebert-Uphoff, I., Gagne, D. J. II, & Bostrom, A. (2021). The need for ethical, responsible, and trustworthy artificial intelligence for environmental sciences. *arXiv preprint arXiv:2112.08453.*
33. Miloshevich, G., Cozian, B., Abry, P., Borgnat, P., & Bouchet, F. (2023). Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data. *Physical Review Fluids*, *8*(4), 040501.
34. Naveau, P., Hannart, A., & Ribes, A. (2020). Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and Its Application*, *7*(1), 89–110.
35. Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., ... & Matias, Y. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563.
36. Olivetti, L., & Messori, G. (2024). Advances and prospects of deep learning for medium-range extreme weather forecasting. *Geoscientific Model Development*, *17*(6), 2347–2358.
37. Otto, F. E. (2023). Attribution of extreme events to climate change. *Annual Review of Environment and Resources*, *48*(1), 813–828.
38. Pasche, O. C., Wider, J., Zhang, Z., Zscheischler, J., & Engelke, S. (2025). Validating deep learning weather forecast models on recent high-impact extreme events. *Artificial Intelligence for the Earth Systems*, *4*(1), e240033.
39. Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., ... & Anandkumar, A. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214.*
40. Prabhat, Kashinath, K., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., ... & Collins, W. (2020). ClimateNet: An expert-labelled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development Discussions*, *2020*, 1–28.
41. Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russell, T., ... & Sha, F. (2024). WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, *16*(6), e2023MS004019.
42. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.
43. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... & Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, *55*(2), 1–96.
44. Ronco, M., Tárraga, J. M., Muñoz, J., Piles, M., Marco, E. S., Wang, Q., ... & Camps-Valls, G. (2023). Exploring interactions between socioeconomic context and natural hazards on human population displacement. *Nature Communications*, *14*(1), 8004.
45. Runge, J. (2020). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(9), 093121.
46. Runge, J., Gerhardus, A., Varando, G., Eyring, V., & Camps-Valls, G. (2023). Causal inference for time series. *Nature Reviews Earth & Environment*, *4*(7), 487–505.
47. Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G., Del Ser, J., Casillas-Pérez, D., Kadow, C., ... & Castelletti, A. (2024). Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: A review. *Theoretical and Applied Climatology*, *155*(1), 1–44.
48. Schlund, M., Eyring, V., Camps-Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M. (2020). Constraining uncertainty in projected gross primary production with machine learning. *Journal of Geophysical Research: Biogeosciences*, *125*(11), e2019JG005619.
49. Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., ... & Yuval, J. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200097.
50. Sun, S., Zhang, H., & Xu, J. (2022). Multimodal data fusion for climate and environmental analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(9), 4362–4379.
51. Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2020MS002258.

52. Trok, J. T., Barnes, E. A., Davenport, F. V., & Diffenbaugh, N. S. (2024). Machine learning–based extreme event attribution. *Science Advances*, *10*(34), eadl3242.
53. Tuia, D., Schindler, K., Demir, B., Zhu, X. X., Kochupillai, M., Džeroski, S., ... & Camps-Valls, G. (2024). Artificial intelligence to advance Earth observation: A review of models, recent trends, and pathways forward. *IEEE Geoscience and Remote Sensing Magazine*, *12*(3), 8–37.
54. Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating physics-based modeling with machine learning: A survey. *ACM Computing Surveys*, *55*(4), 1–34.
55. Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., & Shen, C. (2021). Physics-guided deep learning for rainfall–runoff modeling by considering extreme events and monotonic relationships. *Journal of Hydrology*, *603*, 127043.
56. Yang, R., Hu, J., Li, Z., Mu, J., Yu, T., Xia, J., ... & Xiong, H. (2024). Interpretable machine learning for weather and climate prediction: A review. *Atmospheric Environment*, *338*, 120797.
57. Zhang, M., Fernandez-Torres, M. A., & Camps-Valls, G. (2024). Domain knowledge-driven variational recurrent networks for drought monitoring. *Remote Sensing of Environment*, *311*, 114252.