Award Winner

Forecasting What Has Never Happened

Long (Mark) Nguyen, ASA, ACAS and Stefanos Orfanos, FSA, Ph.D.

This essay examines whether AI weather models can reliably forecast the high-impact extremes that drive catastrophe losses as climate change pushes conditions beyond the historical record. We present evidence from recent tropical cyclone experiments and European windstorm case studies showing that AI models often smooth peak hazard intensities—the very values that matter most for solvency. The limitation is not a bug to be fixed, but a consequence of how these models learn: they are empirical interpolators, not physical extrapolators. We propose FACT (Forecast with AI; Anchor to physics; Calibrate to exposure; Tie to governance) as a hybrid workflow that leverages AI's speed for scenario exploration while using physics-based models to adjust tail behavior.

INTRODUCTION: WHEN EXPERIENCE FALLS SHORT

Hurricane Ian in 2022 made landfall in Florida as a Category 4 storm, bringing catastrophic storm surge and up to 27 inches of rainfall, causing an estimated $113 billion in damage. That same year, monsoon rains and glacier melt left one-third of Pakistan underwater, affecting 33 million people and causing over $30 billion in total losses. In 2023, Canada saw its worst wildfire season on record, with more than 18 million hectares burned—about seven times the national 10-year average—driving prolonged smoke across parts of the United States.

These events share an uncomfortable theme: they sit outside what these communities have experienced before. For actuaries, they represent exactly the outcomes we need to anticipate and the ones we struggle with the most. Our traditional tools like credibility theory, experience rating, and statistical frequency analysis, are fundamentally backward-looking. They assume the future will resemble a weighted average of the past. Climate science tells us this assumption is breaking down.

At the same time, weather prediction is undergoing one of the fastest shifts in its history. Since around 2022, AI weather models such as Huawei's Pangu-Weather and Google DeepMind's GraphCast have demonstrated forecast skill comparable to or exceeding traditional physics-based numerical weather prediction (NWP) systems, producing accurate global predictions at a fraction of the computational cost.

But underneath the hype lies a crucial question: can AI reliably forecast rare, high-impact events? The answer, increasingly supported by research, is "not yet—at least not alone." And that limitation matters more as climate change pushes weather conditions into territory the historical record never prepared us for.

TWO WAYS OF KNOWING THE ATMOSPHERE

To understand why AI models struggle at the tails and why that issue is structural, we need to distinguish two fundamentally different approaches to forecasting. This is not just a technical detail; it is the conceptual foundation for everything that follows.

THE PHYSICIST'S WAY: EXTRAPOLATION FROM FIRST PRINCIPLES

Traditional NWP starts with physics. Air flows from high to low pressure. Water changes phases according to thermodynamics. Earth's rotation bends winds through the Coriolis effect. The atmosphere can be represented as a three-dimensional grid, and partial differential equations—conservation of mass, momentum, and energy—are solved forward in time.

Physics-based models have a critical advantage: they can extrapolate. They can simulate a Category 5 hurricane even if the historical record contains few storms at that strength. If the initial conditions support such a storm, the equations will produce it. Nothing in the mathematics forbids outcomes that have never been observed, like 200-knot winds or 60°C surface temperatures, if physics permits them. But this power comes at a cost: each forecast cycle can take hours on dedicated supercomputers, and running many scenarios at high resolution is often impractical when decisions are time-sensitive.

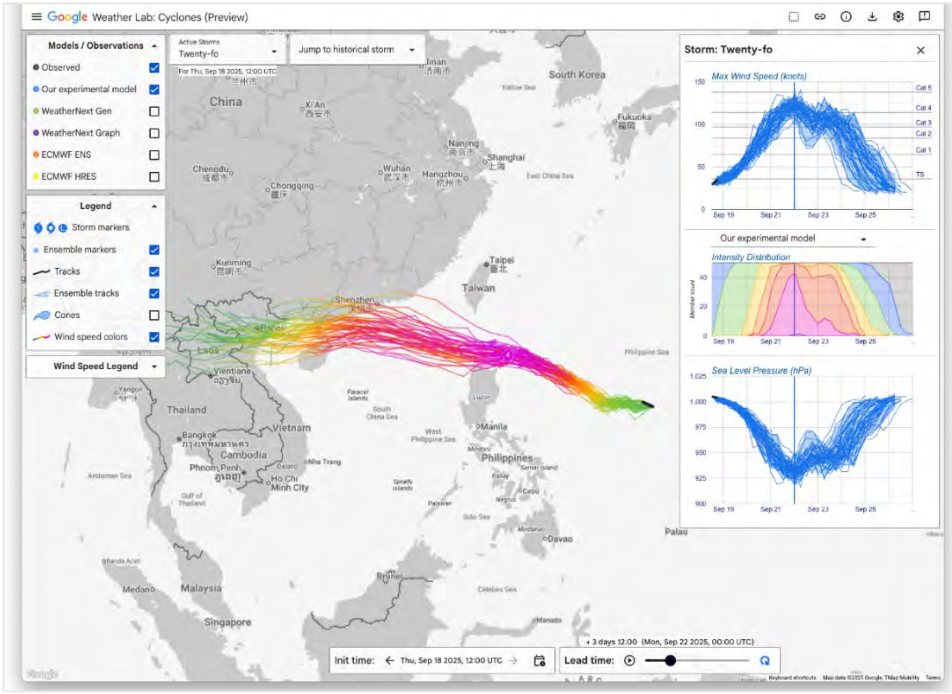
THE DATA SCIENTIST'S WAY: INTERPOLATION FROM PATTERNS

AI weather models take a different approach. Instead of solving equations, they learn from examples. Feed decades of global weather data (typically the European Centre for Medium-Range Weather Forecasts reanalysis v5 (ERA5), which reconstructs atmospheric states from 1979 to the present) to a neural network and ask it to learn what tomorrow tends to look like given today. The "intelligence" is pattern recognition at scale, identifying statistical regularities across four decades of weather.

These models interpolate. They encode correlations: when the atmosphere looked like this in the past, it subsequently looked like that. A model trained on data where maximum recorded winds reached 180 knots has no principled basis for predicting 200-knot winds. It may do so accidentally, but the prediction carries no theoretical warrant. More often, the model regresses toward familiar territory; confronted with inputs unlike anything it has seen, it produces outputs closer to the training mean.

Figure 1

EXAMPLE OF RAPID, ENSEMBLE-STYLE CYCLONE FORECASTING



This example is an AI-driven interface, showing many plausible tracks and intensity trajectories from a single initialization. Source: Google Weather Lab, Cyclones preview (https://developers.google.com/weathernext/guides/research). Licensed under CC BY 4.0. No changes made.

WHY THIS DISTINCTION MATTERS FOR CAPITAL

Actuaries will recognize this tension. Credibility theory is fundamentally interpolative: it weights recent experience against a broader mean, assuming the future looks like a weighted average of the past. Catastrophe modeling is fundamentally extrapolative: it uses physical principles (engineering, seismology, atmospheric dynamics) to simulate events that have never happened.

AI is the ultimate credibility engine: fast, data-driven, excellent at the middle of the distribution. Physics is the ultimate catastrophe engine: slower, more expensive, but capable of reaching into the tails. The question for actuaries is not which to choose, but how to allocate each to its comparative advantage.

Several independent studies now document the same pattern: machine learning weather models match or exceed physics-based forecasts for typical conditions but systematically underestimate rare extremes.

CONTROLLED EXPERIMENTS: REMOVING THE TAILS

Sun and collaborators (2025) ran controlled experiments with NVIDIA's FourCastNet, deliberately removing strong tropical cyclones from training data. When all Category 3–5 storms were excluded globally, the model failed to forecast Category 5 storms. When strong storms were removed only from the Atlantic but retained in the Pacific, Atlantic Category 5 performance recovered substantially because the model transferred patterns learned elsewhere.

The researchers call this distinction translocation vs. extrapolation. Machine learning models can recognize a pattern seen in one region when it appears in another—they can translocate. They struggle to predict intensities beyond anything in the global training set; hence they cannot truly extrapolate.

CASE STUDY: DUBAI 2024

Dubai's April 2024 rainfall illustrates the power of translocation. Some locations received over 250mm in 24 hours—roughly double anything in the regional historical record. This was a "gray swan" for the Arabian Peninsula: unprecedented locally, but dynamically similar to mesoscale convective systems the model had seen thousands of times in the tropics and the U.S. Midwest.

GraphCast predicted the event eight days in advance, demonstrating that AI can help identify risks in regions with poor local historical data, provided the hazard mechanism is not unique to that location. For actuaries, this is good news: AI can scan for "gray swans" that local experience would miss. But even here, the model underestimated peak intensity. Translocation worked; extrapolation did not.

CASE STUDY: STORM CIARÁN 2023

In November 2023, Storm Ciarán—a rapidly intensifying "bomb" cyclone—struck Western Europe. Charlton-Perez and colleagues compared four leading AI models (FourCastNet, FourCastNet-v2, Pangu-Weather, and GraphCast) with ECMWF's physics-based forecasts.

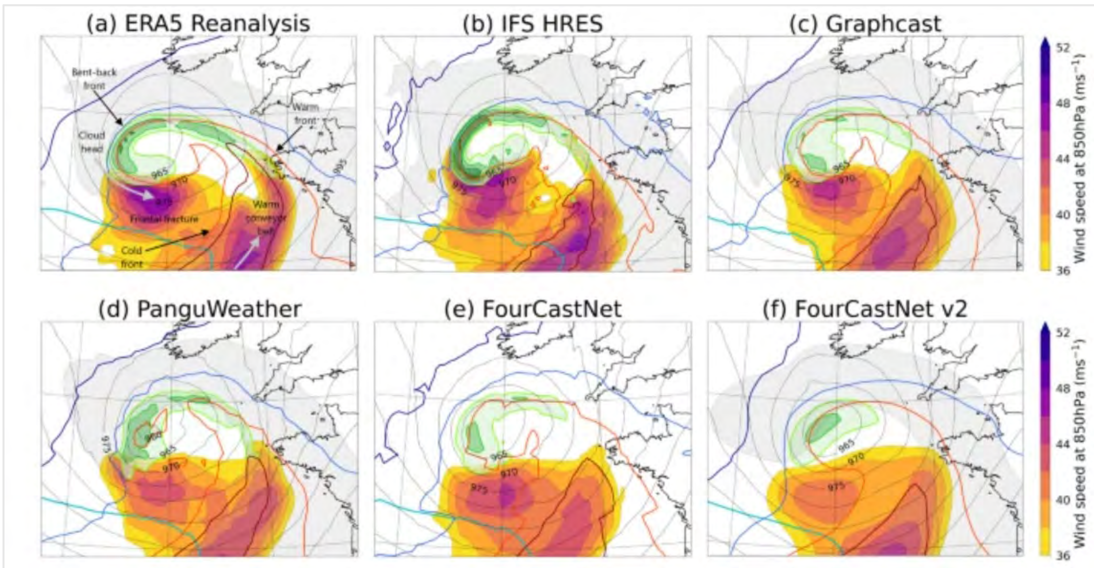
The AI systems captured the large-scale structure, timing, and rapid deepening of the cyclone. But all of them systematically underestimated the strongest near-surface winds. They got the storm at roughly the right time and shape but missed the peak gusts that drive losses.

The underlying mechanism relies on mathematics. Most AI weather models minimize mean squared error (MSE), which penalizes large mistakes heavily. The loss-minimizing strategy is to produce "safe" predictions near the middle of the distribution—to smooth.

THE FINANCIAL TRANSLATION: SMALL ERRORS, LARGE LOSSES

Why should actuaries care about a 10% wind speed error? Because wind damage scales roughly with the cube of gust speed above a threshold. If the AI predicts 45 m/s gusts but the reality is 50 m/s (a 10% miss), the modeled energy and potential damage is understated by approximately 27%. A 15% wind underestimate can push loss error toward 50%.

Figure 2
DYNAMICAL STRUCTURE OF STORM CIARÁN AT 18 UTC ON 1 NOVEMBER 2023



The first two panels show 10-m wind and sea-level pressure from (a) reanalysis data and (b) NWP model, while panels (c)–(f) show the corresponding forecasts from four AI weather models. The AI forecasts exhibit noticeably smoother wind and pressure gradients than the analysis and NWP fields, with weaker peak values in the storm core.
Source: Charlton-Perez, C., et al. (2024). 'AI Weather Prediction Models and Extreme Events: A Storm Ciarán Case Study.' npj Climate and Atmospheric Science. https://doi.org/10.1038/s41612-024-00638-w. Licensed under CC BY 4.0. No changes made.

For a portfolio heavily exposed to European windstorms, relying solely on AI forecasts for capital setting (one-in-200-year return period) would lead to systematic under-capitalization. The "safe" mean prediction of the AI becomes a dangerous underestimation of capital need. Small errors in the meteorological tails become solvency-sized mistakes.

FACT: A HYBRID WORKFLOW

None of this means AI has failed. It means AI and physics need to work together, especially for high-stakes rare events. We propose FACT as a practical workflow for actuaries and risk managers who want to integrate weather forecasts into catastrophe modeling and capital adequacy.

F—FORECAST WITH AI

Define the peril and time horizon relevant to the decision: hurricane season in the Gulf, winter windstorm in Europe, monsoon flooding in South Asia. Rather than using machine learning, use generative AI like GenCast. These models utilize diffusion processes to generate ensembles that preserve variance and fine-scale structure.

Generate hundreds or thousands of plausible atmospheric trajectories by perturbing initial conditions or sampling from the generative model. This takes minutes on GPUs compared to hours for a 50-member NWP ensemble. Translate scenarios into preliminary hazard footprints and overlay on portfolio exposure to identify attention events: scenarios where hazard and exposure coincide in ways that could produce material losses.

This phase exploits AI's comparative advantage: speed. The output is a ranked shortlist, not a final loss estimate.

A—ANCHOR TO PHYSICS

For attention events identified in Phase F, apply physics-based methods to refine tail behavior. Run flagged scenarios through high-resolution NWP, regional dynamical downscaling, or vendor hazard modules calibrated on extreme events. The objective is to sharpen peak winds, resolve localized rainfall maxima, and capture storm surge and flood depths that depend on fine-scale topography.

Document adjustments explicitly: which hazard variables were refined, what methods were used, how upper quantiles changed. This transparency supports model validation and regulatory review. The AI tells you where to look; physics tells you how bad it could get.

C—CALIBRATE TO EXPOSURE

Risk is not just hazard; it is hazard × vulnerability × exposure. Pass the tail-adjusted hazard footprints through vulnerability functions and overlay on your portfolio's specific locations, construction types, and insured values. Look for "translocation hotspots"—areas where the AI predicts a hazard type (e.g., convective flood, rapid cyclone intensification) that local historical experience would miss.

Next, compute gross and net losses under current policy terms and reinsurance structures. Use the scenario set for stress testing: What is the one-in-100-year loss under current exposures? How does the answer change under climate-driven hazard intensification? For longer planning horizons, such as the Own Risk and Solvency Assessment (ORSA) and strategic capital planning, extend the analysis to alternate climate states (+1.5°C or +2°C warming).

T—TIE TO GOVERNANCE

The "black box" nature of AI creates model risk that must be managed through governance, not just mathematics. Connect analysis to action and accountability: use scenario-derived estimates to set risk appetites, adjust pricing, and identify accumulation hotspots requiring limits or exclusions.

Integrate FACT into the ORSA by explicitly documenting the hybrid nature of the assumptions: "Frequency and track derived from AI ensembles; intensity anchored by physics-based models." This satisfies requirements for understanding model components and demonstrates robustness against climate shifts. Make sure to back-test AI hazard estimates against observed events and monitor for drift as climate shifts distributions. Finally, report limitations transparently to boards and regulators: AI models are reliable for the body of the distribution, less so for the tails that drive solvency risk.

WHAT'S NEXT: AI WEATHER ON THE HORIZON

The current generation of ML weather models represents the first wave. Several developments suggest how the field may evolve:

Foundation models are emerging for weather and climate. Microsoft's Aurora, trained on over a million hours of diverse atmospheric and ocean data, aims to generalize across forecasting tasks rather than optimizing for a single prediction type. Early results suggest improved performance on some out-of-distribution events, though rigorous evaluation on true record-breakers remains limited.

Physics-informed architecture embeds conservation laws and dynamical constraints directly into network structure. NowcastNet, for precipitation nowcasting, incorporates advection physics into its design, producing sharper, more physically plausible rainfall predictions than pure pattern-matching approaches.

Generative models like GenCast already produce probabilistic forecasts by sampling from learned distributions. Future systems may better capture tail behavior through loss functions that explicitly weight rare events, or through training augmented with physics-generated synthetic extremes—scenarios simulated by NWP that push beyond the historical record.

None of these developments eliminates the fundamental constraint: empirical models learn from data, and data cannot contain what has not yet occurred. But they suggest a trajectory where AI weather systems become less brittle at distribution edges and better at translocation; more respectful of physical constraints and more honest about what they do not know.

CONCLUSION: THE ACTUARY'S ROLE

Weather forecasting has gained powerful new tools. Deep neural networks can match decades of physics-based development on standard metrics, producing global forecasts in minutes. For ordinary weather, these systems have earned a central role.

But the events that reshape balance sheets are not ordinary—the hurricanes that intensify beyond expectation, the heatwaves that exceed historical records, the floods that surpass engineering tolerances. For these, AI models show consistent weakness: they underestimate intensity, smooth peaks, and regress toward historical averages. The reasons are structural, rooted in training objectives that penalize boldness and data that cannot contain what has not yet occurred.

Physics-based models remain essential for the tails. They encode causal knowledge, such as conservation laws, thermodynamics, and fluid dynamics, which extend beyond any training set.

The path forward is integration. Use AI for fast, cheap exploration of scenario space, identifying where translocation reveals "gray swans" that local experience would miss. Use physics to anchor the extremes, ensuring that the tails of your loss distribution reflect thermodynamic possibility, not just statistical history. Build workflows, like FACT, which allocate each method to its comparative advantage. And develop a risk management framework that is both agile enough for the chaotic future and robust enough for the balance sheet.

REFERENCES

American Academy of Actuaries. 2025. *Climate Scenario Considerations for U.S. Own Risk and Solvency Assessment (ORSA).*

Bodnar, C., W. P. Bruinsma, A. Lucic, et al. 2025. “A Foundation Model for the Earth System.” *Nature* 641: 1180–1187.

Bucci, L., L. Alaka, A. Hagen, S. Delgado, and J. Beven. 2023. *Tropical Cyclone Report: Hurricane Ian (AL092022).* National Hurricane Center.

Charlton-Perez, C., et al. 2024. “AI Weather Prediction Models and Extreme Events: A Storm Ciarán Case Study.” *npj Climate and Atmospheric Science.* https://doi.org/10.1038/s41612-024-00638-w.

Government of Pakistan, Asian Development Bank, European Union, and United Nations. 2022. *Pakistan Floods 2022: Post-Disaster Needs Assessment (PDNA) – Main Report.*

Klawa, M., and U. Ulbrich. 2003. "A Model for the Estimation of Storm Losses and the Identification of Severe Winter Storms in Germany." *Natural Hazards and Earth System Sciences* 3: 725–732.

Price, I., A. Sanchez-Gonzalez, F. Alet, et al. 2025. "Probabilistic Weather Forecasting with Machine Learning." *Nature* 637: 84–90.

Sun, Y. Q., P. Hassanzadeh, T. A. Shaw, and H. A. Pahlavan. 2025. "Predicting Beyond Training Data via Extrapolation versus Translocation: AI Weather Models and Dubai's Unprecedented 2024 Rainfall." *arXiv* preprint arXiv:2505.10241.

Sun, Y. Q., P. Hassanzadeh, M. Zand, et al. 2025. "Can AI Weather Models Predict Out-of-Distribution Gray Swan Tropical Cyclones?" *Proceedings of the National Academy of Sciences* 122 (21): e2420914122.

Zhang, Z., E. M. Fischer, J. Zscheischler, and S. Engelke. 2025. "Numerical Models Outperform AI Weather Forecasts of Record-Breaking Extremes." *arXiv* preprint arXiv:2508.15724.

* * * * *

Long (Mark) Nguyen, ASA, ACAS, is a graduate student at Georgia State University. He can be reached at tuanlongnguyenha@gmail.com.

Stefanos Orfanos, FSA, Ph.D. is a Clinical Assistant Professor at Georgia State University. He can be reached at sorfanos@gsu.edu.