Hot Topics:

Machine Learning for Weather Forecasting & Heatwave Prediction to Medium Range & Beyond

A Guide for Meteorologists – Burn After Reading

Why read this report?

Written for operational or research meteorologists everywhere, this is a guide on how machine learning is being used in weather prediction, with a focus on the challenge of extreme heat. Cutting through the hype – it's an approachable review on what is happening and what to watch. If you are wondering where things are heading, keep reading – some of the fieriest issues are still unresolved.

29th May 2025

Vicky Lucas

1. Degrees of change: introduction

Machine learning (ML) is reshaping meteorological modelling, offering new ways to emulate, improve or replace a range of weather and climate modelling (Chantry et al., 2021). In just a few years techniques have moved from focused applications, such as radar nowcasting, to full global models using machine learning for weather prediction (MLWP) that rival, and in some cases outperform, traditional numerical weather prediction (NWP) systems. The shift is not just in who builds these models, often technology companies, but in how forecasts are produced. MLWP does not solve physical equations, but learns patterns in data, based on techniques used in other machine learning applications such as recommendation engines, image generation and Large Language Models (LLMs).

Many MLWP systems are trained on historical and constructed weather datasets, very often reanalysis – and especially ERA5 (Climate Data Store). MLWP model training requires significant data, time and energy, but once trained, operational ML models can produce forecasts in seconds or minutes on minimal hardware compared with NWP supercomputers. This speed opens up possibilities for larger ensembles, rapid updates, and hybrid workflows.

Something to bear in mind is that this is a domain in flux, no architecture clearly prevails. Groups are experimenting with configurations, progressively leapfrogging their own and others' achievements. While many MLWP models appear in peer-reviewed journals, some remain as pre-prints – including references in this report (e.g., arXiv), and some make open-source versions available.

As developments continue, the plethora of global general MLWP models can be divided by their training regimes (Shi et al., 2025 – arXiv):

- **Deterministic predictive learning**, as used in *FourCastNet*, uses supervised training – labelled data where each input comes with a corresponding output – a learnt relationship. These models minimise forecast error, are fast to run, but accumulate errors over longer lead times.
- **Probabilistic generative modelling**, such as *GenCast*, uses diffusion to generate ensemble forecasts, offering uncertainty modelling without explicit perturbations of NWP ensembles. Generative models learn the distribution of the training data, then generates new samples.
- **Foundation models**, such as *Aurora*, pre-trained on extremely large and diverse datasets to fulfill a broad range of tasks, transferring learning to other datasets either with fine-tuning on new data or prompting e.g., from weather to air quality modelling.

This broad training delineation is a framework to understand ML in meteorology. In some instances, deep learning, a subset of ML, would be a more accurate term, but ML is used here for simplicity. A glossary of ML terms is included at the end of this report – when new terms are used a hyperlink to the glossary is provided.

This report follows how MLWP is shifting from research to real-world use. It starts by tracing the recent proliferation of global ML models, introducing some of the notable participants and a little on how they work. Next, is an exploration on what is happening in forecasting heatwaves and extending the forecasting window beyond the medium range and into sub-seasonal to seasonal (S2S). The final sections look ahead to where the field is going – concluding with unresolved tensions and challenges.

2. Mercury rising: chronology of machine learning weather models

The rapid development of general global machine learning weather models reflects technology company efforts in meteorological modelling and capitalising on techniques, such as used in Large Language Models e.g., ChatGPT (transformers) or image generation (diffusion). Since 2022 MLWP models have made headlines on outperforming respected operational NWP. These ML models were optimised for minimising average errors, but later models have switched to generative modelling and two types of this approach, *GenCast* and *Aurora*, are discussed in more detail later.

A timeline of general global ML models 2022 to 2025 is shown in Figure 1, and a glossary of ML terms used is provided at the end of this report.

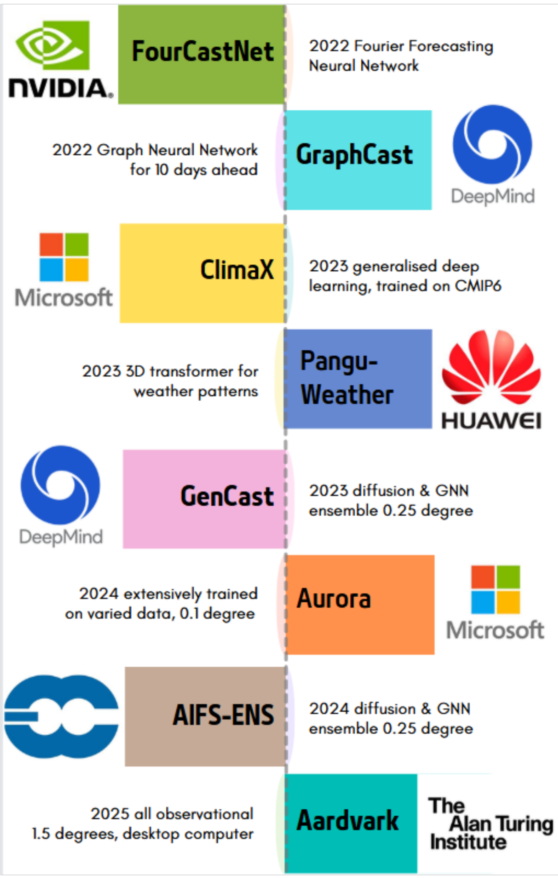

Figure 1: timeline of selected ML weather models

FourCastNet (Pathak et al., 2022 – arXiv) was the first general-purpose deterministic MLWP model based on a neural network that rivaled the accuracy of operational NWP. *FourCastNet* was trained on 10TB of data, emulating atmospheric dynamics learnt from ERA5 reanalysis. It can be run globally in seconds at 0.25° resolution. The breakthrough was twofold. Firstly, using the Fourier Neural Operator approach (Li et al., 2021 – arXiv) which learns mapping between entire spatial fields (function space) rather than individual and neighbouring grid points. Its Fourier architecture captures both low- and high-frequency components, efficiently representing global and local structures. Second, the integration of a Vision Transformer enables the model to learn long-range spatial dependencies. The Vision Transformer was a landmark paper on self-attention, which allows networks to prioritize relevant information when learning; published by Google in 2017 and used in LLMs such as ChatGPT – it is one of the most cited papers this century (Pearson et al., 2025). Together, these innovations enabled fast learning, high resolution, and efficient computation.

GraphCast (Lam et al., 2023) appeared in pre-prints soon after, later published in *Science*, applying Graph Neural Networks to global forecasting. GNNs represent the atmosphere as an icosahedral mesh, enabling uniform spatial resolution. Nodes (grid points) are connected by edges which define the flow of information between all nodes – allowing the model to learn teleconnections regardless of distance or input grid regularity. Forecasting over 1,000 variables up to 10 days ahead, *GraphCast* achieved greater accuracy than ECMWF IFS conventional NWP model – including at 10 day lead times for the top 2% of hot days over land (Lam et al., 2023).

Previous models had taken their names from their architecture, but *Pangu-Weather* changed that – named after a primordial mythical being who separated heaven and earth and became geographic features. This deterministic model introduced two new approaches (Bi et al., 2023). Firstly, it used a three-dimensional approach, using vertical structure via pressure level data. Secondly it trained separate models for 1, 3, 6 and 24-hour lead times, which reduced errors over long lead times.

Throughout these advances ECMWF, the European Centre for Medium-Range Weather Forecasts – both a research institute and operational service – has been central in the application of ML. The ECMWF operational NWP, the IFS HRES, is often used as the benchmark model comparison, or the ensemble IFS ENS. ECMWF developed its own deterministic graph-based (GNN) system, *AIFS* (Lang et al., 2024 – arXiv), and an ensemble *AIFS-ENS* based on diffusion (Figure 1) (Lang et al., 2024). Diffusion is the same approach used for image generation and in a later MLWP model, *GenCast* – features in the next section of this report, where diffusion will be discussed in more detail. The *AIFS-ENS* was made operational in early 2025, alongside the longstanding conventional NWP the IFS.

ECMWF displays third-party MLWP outputs in real-time and evaluates model performance. Figure 2 shows comparative performance through the root mean squared error of two-metre air temperature for the northern extra-tropics, for winter 2024/25 – showing that *Aurora* consistently achieves the greatest accuracy on that broad metric – outperforming the operational NWP, IFS. A range of charts and performance scores are available from the ECMWF chart catalogue.

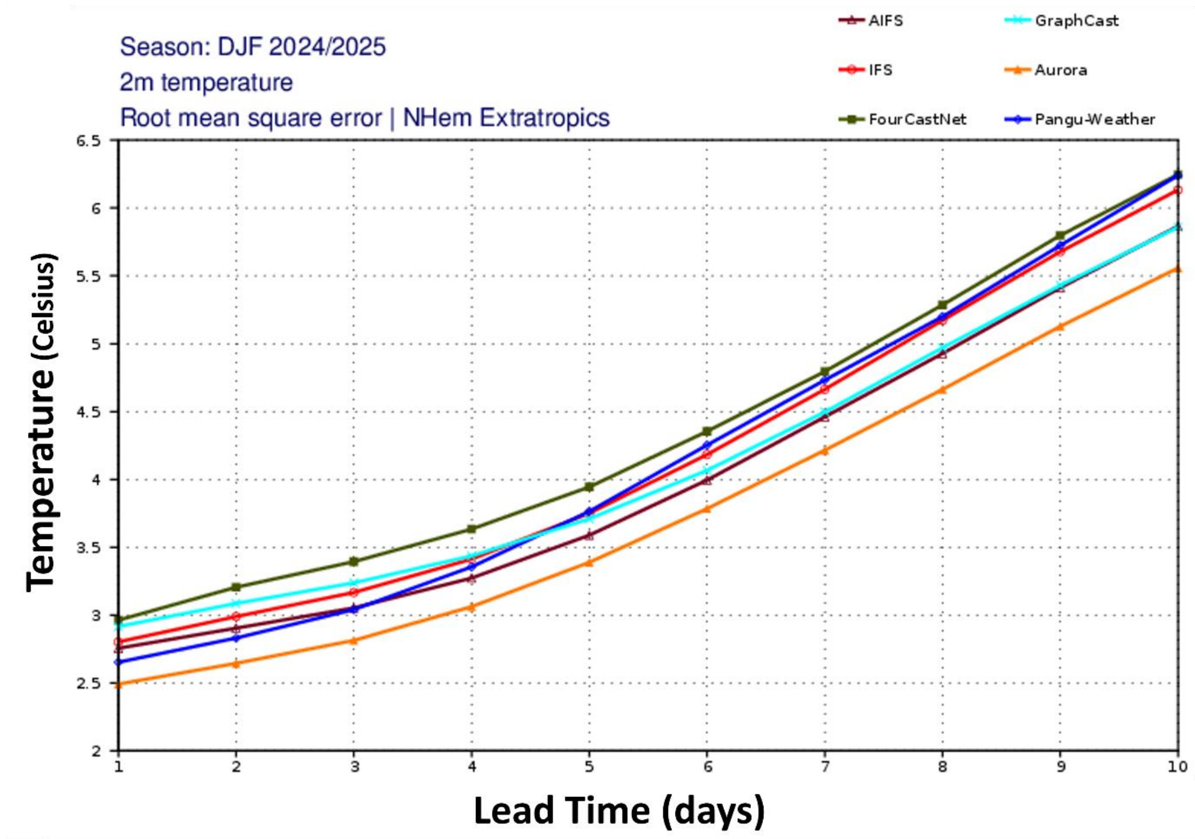

Figure 2: RMSE 2m temperature, winter 24/25 northern extra-tropics - consistently best performance from Aurora

IFS (ECWMF operational NWP) (red) and five MLWP models: AIFS (ECMWF) (brown), FourCastNet (green), GraphCast (cyan), Pangu-Weather (dark blue) & Aurora (orange) – available from ECMWF

ClimaX (Nguyen et al., 2023 – arXiv) was the first foundation model, trained on truly vast and heterogeneous datasets including ERA5 and the climate model output of CMIP6. The approach introduced a shared transformer encoder pre-trained, and which can be fine-tuned to more specific tasks. *ClimaX* is capable of global and regional forecasts, also sub-seasonal and climate projections, and downscaling. Similarly, *Prithvi WxC* (Schmude et al., 2024 – arXiv) is a later foundation model, a collaboration between IBM and NASA.

Another first, *Aardvark*, has recently been published in *Nature* (Allen et al., 2025). The system has two notable features, firstly it can run on a desktop computer in minutes (Turner, 2025). Secondly, whilst it has been trained on ERA5, it does not need any NWP products to run, it only requires observations as input such as satellite and weather station data. This end-to-end system, fully observation-driven, has been an aspiration in the field (McNally et al., 2024). *Aardvark* shows strong performance against operational NWP, even with a coarse 1.5 degree output and using just 10% of inputs to existing NWP systems (Turner, 2025). These features make *Aardvark* valuable for applications like disaster preparedness in low-resource settings.

The details: GenCast & Aurora

To illustrate recent architectures, two models are detailed here, *GenCast* and *Aurora* (Figure 1). *GenCast* is a probabilistic generative model for medium range 15 day forecasting (Price et al., 2025) while *Aurora* is a foundation model which can be tuned to multiple applications including forecasting air quality and tropical cyclone tracks (Bodnar et al., 2024 – arXiv).

GenCast is a generative model – in training it learns the underlying data distribution and then produces new realistic samples. This generative function comes from using a diffusion process, based on noise. In training, to predict conditions 12 hours ahead (T+12), *GenCast* takes two input fields from ERA5, effectively T-12 and T+0. Rather than forecasting T+12 directly, it computes the residual (i.e., the change between T–12 and T+0) and adds scaled noise to create the first guess. *GenCast* then iteratively removes noise over multiple steps. By perturbing the tendency rather than the absolute state, it preserves coherent atmospheric structures. *GenCast* was pre-trained at 1° resolution on 40 years of ERA5 data, then fine-tuned at 0.25°.

In forecast mode, known as inference in ML, *GenCast* predicts each T+12 using the two most recent states, continuing until it reaches T+360 (15 days). This type of rolling prediction is called autoregression in ML. Generating an ensemble is straightforward due to the diffusion approach; each ensemble member starts from a different scaled noise sample. 50 ensemble members were chosen to match ECMWF IFS ENS – a lot more could be produced. While diffusion models are computationally intensive—each step involves multiple denoising passes—*GenCast* can produce a full 15-day forecast in minutes per member. Its Graph Neural Network (GNN), adapted from *GraphCast* (discussed earlier), captures and supports spatial dependencies across time steps. *GenCast* outperformed IFS ENS, including extreme temperatures, the 99.9th and 99.99th percentiles, out to 15 days (Price et al., 2025).

In contrast*, Aurora* is a foundation model – the term foundation model came from Stanford in 2021 – where the model is a common basis for many task specific adaptations. *Aurora* is pre-trained on more than one million hours of global Earth system data, including weather reanalyses, operational forecasts, and the CMIP6 climate simulations. It contains 1.3 billion parameters and uses a 3D Swin Transformer, an architecture adapted from computer vision. The Swin, shifted window, transformer improves efficiency by dividing data into local windows and then shifting them between layers to capture broader spatial and vertical dependencies. *Aurora* also employs perceiver-based encoders and decoders which allow the model to ingest varying types and shapes of data, transforming them to a common internal format to combine different input data.

These architectural choices allow Aurora to model complex Earth system dynamics across domains. After pretraining, it can be fine-tuned for tasks like 10-day global weather forecasting, 5-day air quality prediction, ocean wave modelling, and tropical cyclone tracking. *Aurora* (Bodnar et al., 2024 – arXiv) produces 10-day global forecasts in under a minute and 5-day air quality predictions in seconds.

Summary

From the fast run times of *FourCastNet* to the flexible tuning of *Aurora* and the ensemble uncertainty quantification of *GenCast*, MLWP is approaching, and by some metrics exceeding NWP performance, all with significantly reduced operational computational cost. The enablers of that progress include:

- Open access to high-resolution datasets such as ERA5,
- Cross-disciplinary architecture design (e.g. transformers, diffusion),
- Collaboration between research institutes, operational centres, and tech companies,
- Access to MLWP on GitHub and Hugging Face – for sharing data, code and documentation.

For further reading, Chen et al. (2023) offers a good overview of methods and architectures, while Zhang et al. (2025) provides an up-to-date review of knowledge, methods, trends and challenges. Waqas et al. (2024) offers a systematic review of AI integrated with NWP. For more on MLWP comparative performance, Google Research has a scorecard of several models at WeatherBench2.

3. On the horizon: ML in heatwave forecasting & beyond medium range

General MLWP at medium range & S2S

MLWP is extending skill into the medium and sub-seasonal range (15–46 days). Two deterministic general purpose transformer models, *FuXi* and *FengWu* have moved to architectures with ensemble generation for forecasts out to six weeks in S2S versions. *FuXi-S2S* demonstrates improved skill over ECMWF IFS S2S for several metrics, including 2m temperature and Madden-Julian Oscillation (MJO) prediction up to day 36 (Chen et al., 2024). *FengWu-W2S* uses ocean–atmosphere–land coupling constraints for physical consistency (Ling et al., 2024). The results for 2m temperature show comparable but greater skill than ECMWF S2S and *FuXi-S2S* at three to six weeks. *FengWu-W2S* can predict the MJO to 37 days, slightly longer than ECMWF S2S or *FuXi-S2S*, and predicts the North Atlantic Oscillation with greater skill than ECMWF S2S at four to six weeks.

Climate

ML models are also being tested in climate applications. *NeuralGCM* (Kochkov et al., 2024) is a hybrid model trained on ERA5 that performs well for 10–15 day forecasts and can simulate climate metrics over multiple decades when forced with prescribed sea surface temperatures. The model is not coupled to land or ocean and does not include greenhouse gases or aerosols. *NeuralGCM* does not show a clear trend of increasing error when initialised further into the future from the training data. 'Our results provide strong evidence for the disputed hypothesis that learning to predict short-term weather is an effective way to tune parameterisations for climate.' (Kochkov et al., 2024).

ML improving local & regional NWP for extreme heat

Machine learning is being used to improve heatwave forecasting at local scales and to extend skill in forecasting surface extremes. At the urban level, ML has been used for bias correction and downscaling. A recent London study (Blunn et al., 2024) used ML to combine high resolution NWP output with citizen science weather station observations during heatwaves. ML reduced mean absolute temperature error by 11%, identifying latent heat flux as the most important predictor of temperature bias. In regional modelling, ML can support sensitivity analysis of NWP physical schemes. One example from Australia (Reddy et al., 2023) used ML to identify, of twenty four parameters, the two key drivers of modelled surface temperature and relative humidity predictions: the shortwave radiation scattering parameter and saturated soil water content multiplier. These two studies show complementary uses of ML in improving conventional NWP.

MLWP extending lead times for extreme heat prediction

Papers from 2022 to 2025 exploring the use of ML approaches for extending the lead times of heatwave predictions is detailed in Table 1, and several discussed below. The table also includes recent papers on explainable AI (XAI) discussed in the next section, and two review papers.

Among the more recent and longer-range examples is the Weirich-Benet et al. (2023) study, which applied both linear regression and random forests to forecast summer heatwaves in Central Europe at lead times of 1–6 weeks. Key predictors included 500 hPa geopotential height, soil moisture, and sea surface temperatures. While model performance declined with increasing lead time, the machine learning outputs outperformed persistence and climatology. Beyond two weeks, their forecasts were as skilful as the ECMWF sub seasonal ensemble-mean hindcast for the region.

Lopez-Gomez et al. (2023) explored ML configurations trained specifically on extreme temperatures — rather than all temperatures. This adjustment led to improved performance over persistence and comparable skill with ECMWF S2S after day 14. The MLWP could forecast out-of-sample events, but struggled with extremely hot days, above the 95th percentile.

In a case study of the North American Pacific Northwest 2021 heatwave, the anomalous temperatures were only predicted 2 to 5 days prior, although at about a month ahead the MLWP was similar in structure to the ECMWF ensemble. Duan et al. (2025), apply the *NeuralGCM*, discussed above, to explore the Pacific Northwest heatwave, which it replicates.

Xie et al. (2024), used a Convolutional Neural Network – an approach that finds local spatial correlations in gridded data – to forecast heatwaves in China up to 30 days ahead. The method filtered out high-frequency signals (<10 days) and isolated the 10–90 day low-frequency background state, using inputs from NCEP/NCAR reanalysis and polar-orbiting satellite data. While forecast skill was lower up to 20 days, the model outperformed both the China Meteorological Administration NWP and ECMWF S2S ensemble mean between days 20 and 30 when averaged across the country.

Table 1: summary of papers on ML approaches applied to heat extremes from 2022 to 2025

Reference & link	***Data or model***	***Area***	***Technique***	***Lead time or purpose***
Ennis et al., 2025	*GraphCast, FuXi, Pangu,* GEFS	USA	MLWP – NWP comparison	Up to 20 days
Lovo et al., 2025	CESM (NCAR)	France	Multiple ML models, also XAI study	XAI
Shafiq et al., 2025	5 years weather observations	Lahore	XAI study, LSTM best performance	1-3 days, XAI
Camps-Valls et al., 2025	Multiple	Global	Review of methods	Review paper
Duan et al., 2025	ERA5	Pacific Northwest	NeuralGCM and ensemble comparison E3SM NWP	6 days
Xie et al., 2024	NCEP/NCAR reanalysis	China	CNN with filtering	Up to 30 days
Weirich Benet et al., 2023	ECMWF	Europe	Linear and Random Forest	1-6 weeks
Lopez Gomez et al., 2023	ERA5	Global	Multiple ML models	Up to 28 days
Salcedo-Sanz et al., 2023	Multiple	Global	Review of methods	Review paper

MLWP for extremes – open questions

Despite encouraging case studies, recent reviews caution that MLWP is not ready for operational forecasting of extremes. Olivetti & Messori (2024) highlight three key limitations: (i) most models are tuned for average forecast skill, not extremes; (ii) architectures are not optimised for the limited data available on rare events; and (iii) assumptions about error distributions are often too simplistic. They also note that leading global MLWP lack validation for extreme event prediction.

Salcedo-Sanz et al. (2023) provide a history and broader critique, focused on extreme events and a literature review of heatwaves. They also emphasise the limited data available for training on rare events and argue that most current ML systems lack integration with physical climate knowledge to achieve reliable predictions. The authors call for the use of multiple reanalysis datasets, improved model transparency through explainable AI (XAI), and greater focus on compound and concurrent extremes. Similarly, the review paper of Camps-Valls et al., (2025) discusses the hurdle of limited training data, and of deploying understandable models – needed for gaining trust. Many authors underscore that MLWP requires consistent and transparent development for extremes.

4. The heat is on: latest developments

As MLWP matures, researchers are responding to critiques — including the validation of extreme event forecasts. Recent work evaluated three global MLWP models (*FourCastNet, Pangu-Weather, GraphCas*t) alongside ECMWF IFS HRES for events including the 2021 Pacific Northwest heatwave (Pasche et al., 2025). While ML models captured the broad structure of the event, they underperformed when spatially and temporally aggregated. *FourCastNet* had the largest errors; *Pangu-Weather* tended to overestimate affected areas for a given threshold. At lead times under a week, the IFS HRES remained the most accurate. Notably, MLWP have few variables – lacking surface relative humidity – and the authors indicated the need for sea surface temperature and soil moisture inputs for longer lead times. A separate study examined 60 U.S. heatwaves and found GraphCast consistently outperformed *Pangu-Weather* and the US GEFS ensemble across most regions with ERA5 used as ground truth (Ennis et al., 2025 – arXiv). The open-source nature of some MLWP models allows for this independent analysis and verification – which is likely to continue – assisted by the WeatherBench2 (Rasp et al., 2023 – arXiv) database of extreme event cases, (also ClimateBench (Watson-Parris et al., 2022)) and ECMWF adding MLWP forecast evaluation to its database of extreme cases (Magnusson, 2023).

Explainability remains a major concern. While studies by Lovo et al. (2025), Shafiq et al. (2025), and Wei et al. (2025 – arXiv) have applied explainable AI (XAI) tools to heat forecasting, these often rely on simplified models or components. As Lovo notes, Convolutional Neural Networks – the earlier predictive approach for MLWP suited to gridded data and used e.g., by GraphCast – 'are black boxes even when using XAI tools'. More transparent methods, hybrid models combining ML with physical constraints, are important for scientific trust and likely to be so for public trust.

A potential trend is the integration of large language models (LLMs) to support forecast interpretation and decision-making. WildfireGPT, developed as a proof-of-concept, combined extreme weather projections, literature, and user prompts to generate geospatially contextualized response guidance (Xie et al., 2024 – arXiv). Designed to be user-centric and visually integrated, it outperformed generic LLMs in ten case studies. A similar 'HeatwaveGPT' could help bridge the gap between technical NWP, ML outputs, operational needs and actions, and public-facing messaging.

Looking ahead, the field is still evolving and diversifying. *Aardvark*, capable of generating forecasts directly from observations, represents a shift in architecture – able to include some measurements too complex for NWP assimilation (McNally et al., 2024). And new systems regularly appear such as the open source WeatherMesh-3 (Du et al., 2025) by an atmospheric sensing company, WindBorne. The S2S timescales and beyond will be a domain of interest; just as MLWP can outperform NWP at short to medium range (1 to 15 days) seasonal predictions will improve for MLWP – including global seasonal or decadal predictions such as *ACE2* (Watt-Meyer et al., 2024, Kent et al., 2025 – arXiv). Another recent development is from NVIDIA – Earth-2 is their commercial digital twin which can be used for kilometre scale weather and climate predictions, it is focused on risk and is a business offering (NVIDIA, 2025). Establishing business models could be a next step for other organisations that have invested resources over the last few years.

As validation continues, explainability evolves and new models appear, the importance of ML in meteorological prediction looks set to remain. Progress could be incremental or radical depending on performance and resourcing pressures – but going too far too fast could undermine trust and acceptance – from science and from the public.

5. Flashpoint: enduring challenges

Machine learning is a growing contributor to meteorological research and, with *AIFS-ENS* at ECMWF, part of the operational suite. Challenges remain – technical, scientific, systemic and social – that will shape the long-term role of ML in meteorological prediction.

i. ***Adoption***: MLWP's headline strengths—global skill scores and fast run times—are useful for benchmarking and publication but operational use is earned through ongoing exposure and case-by-case comparisons. Many MLWP systems still generate overly smooth fields, grow in bias with lead time (Bouallègue et al., 2024), and miss mesoscale features (Bonavita, 2024). MLWP progress builds on decades of NWP based on daily scrutiny and incremental development.
ii. ***Limits***: There are technical and scientific limits to weather prediction. NWP is computationally expensive and time critical operational NWP may be reaching affordable computational limits (Bauer, 2024) – making MLWP attractive operationally if not for the intense training. Rare extremes and data-sparse regions remain a model training issue.
iii. ***Black swans***: Can ML predict extreme events unseen in the training data, compounded by a changing climate? MLWP produces plausible outcomes learnt from past data (Bouallègue et al., 2024). Some studies have shown that MLWP can forecast out-of-sample extreme heat events (Lopez-Gomez et al., 2023). For climate projections, extrapolation to different dynamics or anticipating tipping points may require hybrid models combining ML and scientific knowledge (Patel, 2022 – arXiv).
iv. ***Black boxes***: MLWP is difficult to interpret. Explainability is very important in science-based predictions, for analysis, knowledge discovery and supporting judgements. Explainable AI (XAI) and physics informed neural networks (PINN) are developing fields – but even simple ML models may remain opaque with XAI tools (Lovo et al., 2025). The interpretability gap persists and raised by many, especially for rare extreme events (Zhang et al., 2025, Waqas et al., 2024).
v. ***Trust***: ML may obscure the forecasting process and for public-facing products this could be compounded if LLMs were used in forecast delivery. Public trust depends not only on accuracy, but also accountability. Transparency and examination of prediction performance is necessary to increase public trust in ML (Pasche et al., 2025), as is deploying understandable models (Camps-Valls et al., 2025).
vi. ***Impact***: MLWP must go beyond summary metrics to transparently focus on events and features that matter. Bias in evaluations should be considered, and assessments based on impact to ensure societal equity (Pasche et al., 2025). Evaluations should reflect real-world consequences and consider diverse inputs such as land use or exposure to better guide decisions and address vulnerabilities.
vii. ***Institutions:*** National weather services and research agencies face strategic uncertainty as MLWP challenges traditional NWP frameworks. Bauer (2024) highlights the need for coordinated investments in data infrastructure and high-performance computing to adapt to this shift. Integrating ML into operational systems presents structural and cultural challenges (Zhang et al., 2025), raising questions about whether MLWP will become dominant, operate in parallel with hybrid workflows, or remain subordinate to NWP.

A shift from equation-driven NWP to data-driven MLWP is gaining pace, enabled by decades of atmospheric data, international cooperation and growing cross-sector momentum. The transition is full of tensions – with the flame already lit, but the blaze yet to take hold.

6. Glossary: light relief

Convolutional Neural Network (CNN)	Automatically detects spatial patterns, well suited to gridded data to find local spatial correlation (more from IBM)
Deep Learning	Multilayered neural networks that learn representations from data, a subset of machine learning (more from IBM)
Diffusion model	Generative models, used for image creation, adds noise to the input, then reverses the process to create the output – for sharp structures, extremes and ensemble generation (more from IBM)
Explainable AI	Methods to find links and dependencies in the modelling process, the influence of inputs and sensitivity of outputs (more from IBM)
Fine-tuning (after pre-training)	Adapting a pre-trained model to a specific task using a focused dataset, leveraging existing model knowledge (more from IBM)
Foundation model	Designed to fulfil a broad range of tasks having been pre-trained on immense datasets, able to transfer learning to other tasks by fine-tuning or prompting – such as *Aurora* (more from IBM)
Generative artificial intelligence	Models that learn the underlying distribution of data and that can generate new realistic samples (more from IBM)
Graph Neural Network (GNN)	Designed to work on data connected as nodes and edges, rather than laid out on a regular grid, the model learns relationships between nodes (more from IBM)
Hybrid model	Combining machine learning with physics-based methods to complement, enhance or replace numerical models – keeping core physical constraints – can be used for parameterisation schemes or bias correction. More interpretable than pure ML.
Neural network	A machine learning model that makes decisions by mimicking the way brain neurons work, connecting to others with its own weights and activation thresholds (more from IBM)
Physics Informed Neural Network (PINN)	A type of neural network that embeds physical laws limiting the outputs to being physically consistent – and can be used to make hybrid models (more from Wikipedia)
Predictive artificial intelligence	Discriminative or predictive models focus on classification or regression – examples include *GraphCast* and supervised models such as CNNs, LSTMs and random forests (more from IBM)
Pre-training	Pre-training is in two phases, the model making predictions from inputs and a loss function measuring the difference between the output and the 'correct' answer (more from IBM)
Random forest	A common supervised learning algorithm which combines the outputs of many individual decision trees, the output is either majority vote or averaged (more from IBM)
Supervised learning	Using labelled input-output pairs to make predictions on new data (more from IBM) – see also unsupervised learning
Transformer	A versatile architecture that weights relationships between all parts of an input, captures long-range dependencies and effective for sequential or spatial data (more from IBM)
Unsupervised learning	Finding patterns or groups in data without labelled outputs (more from IBM)

Acknowledgements

This report was produced for the UBC interdisciplinary initiative – the Climate Solutions Research Collective, which annually funds short cross-departmental projects.

Dr Rachel White and Dr Farruk Chishtie proposed the project on Enhancing Heatwave Forecasting and Public Uptake. We have had the benefit of collaborating with Environment and Climate Change Canada, led by Dominique Brunet in the High Impact Weather Research group. The interdisciplinary work has been carried out by Vicky Lucas, lead author on this report, Jing Jang who has been using *GenCast* and *Aurora* and assessing their performance on the North American Pacific Northwest 2021 heatwave, and Emily Rubooga who has analysed interviews with ECCC staff with responsibilities for heat warning dissemination. Dr Ted Scott and Dr Rachel White kindly provided review comments on this document.