Artificial Intelligence in Space Weather Prediction

RAJESH KUMAR MISHRA[1] DIVYANSH MISHRA[3], REKHA AGARWAL[3]
[1]ICFRE-Tropical Forest Research Institute
(Ministry of Environment, Forests & Climate Change, Govt. of India)
P.O. RFRC, Mandla Road, Jabalpur, MP-482021,
INDIA

[2]Department of Artificial Intelligence and Data Science
Jabalpur Engineering College, Jabalpur (MP)
INDIA

[3]Government Science College
Jabalpur, MP,
INDIA

Abstract: - Space weather prediction remains a grand challenge due to the nonlinear, multiscale coupling between solar activity, the heliosphere, and Earth's magnetosphere-ionosphere-thermosphere system. Over the past decade, artificial intelligence (AI) and machine learning (ML) have emerged as powerful complements to physics-based models, offering improved forecast skill, reduced latency, and probabilistic decision support. This review synthesizes recent advances in AI-driven space weather prediction, critically evaluates benchmark performance across forecasting tasks (solar flares, coronal mass ejections, geomagnetic storms, and ionospheric disturbances), and discusses hybrid AI–physics frameworks, explainability, operational readiness, and governance challenges. We argue that AI is reshaping space weather forecasting from empirical pattern recognition toward autonomous, uncertainty-aware prediction systems.

Key-Words: - Space weather, artificial intelligence, machine learning, geomagnetic storms, solar flares, digital twins

Received: April 2, 2026. Revised: May 3, 2026. Accepted: June 21, 2026. Published: July 21, 2026.

1 Introduction

Space weather disturbances originating from solar flares, CMEs, and solar energetic particles—pose systemic risks to satellites, GNSS, aviation, and terrestrial power grids. Forecasting accuracy remains limited by sparse observations, uncertain boundary conditions, and the inherent nonlinearity of Sun–Earth coupling. Traditional physics-based models, though grounded in first principles, often exhibit large forecast errors, particularly for CME arrival time and geomagnetic storm intensity. AI has emerged as a third methodological pillar, alongside theory and numerical simulation. Its strength lies in learning latent relationships across high-dimensional data streams generated by missions such as Solar Dynamics Observatory and operational centers like NOAA Space Weather Prediction Center.

Space weather describes the time-varying physical conditions in the Sun–Earth system driven primarily by solar magnetic activity, including solar flares, coronal mass ejections (CMEs), high-speed solar wind streams, and solar energetic particles. When these phenomena interact with Earth's magnetosphere, ionosphere, and thermosphere, they can induce geomagnetic storms and ionospheric disturbances that adversely affect satellites, global navigation satellite systems (GNSS), radio communications, aviation, pipelines, and electric power transmission networks (Schrijver et al., 2015; Pulkkinen et al., 2017; Baker et al., 2018). With growing societal dependence on space-based and ground-based technological infrastructure, space weather has evolved from a purely scientific concern into a critical issue of economic resilience and national security (Eastwood et al., 2017; Oughton et al., 2017). Predicting space weather, including cosmic ray flux, involves monitoring solar activity, solar wind, and the interplanetary magnetic field. Ground-based observatories, satellites, and space probes play crucial roles in gathering data to understand the dynamics of space weather (Agarwal and Mishra, 2013; Mishra et al., 2024).

Despite decades of progress, reliable space weather prediction remains a grand scientific challenge. The Sun–Earth system is inherently nonlinear and multiscale, spanning spatial scales from solar active regions to the global magnetosphere and temporal scales from seconds to solar cycles. Physics-based models grounded in magnetohydrodynamics (MHD) and empirical coupling functions have provided essential insights into solar wind magnetosphere interactions, yet their operational forecasting skill remains limited by uncertainties in initial and boundary conditions, incomplete physical understanding, and high computational cost (Owens & Forsyth, 2013; Tóth et al., 2005; Riley et al., 2018). In particular, forecasting CME arrival times, geomagnetic storm intensity, and ionospheric response continues to exhibit large errors, especially during extreme events (Manchester et al., 2017; Liemohn et al., 2020).

The rapid expansion of space-based and ground-based observing systems has fundamentally altered the data landscape of space weather research. Continuous high-resolution solar imagery, in situ solar wind measurements, dense GNSS networks, and long-term geomagnetic indices now generate data volumes far exceeding what traditional analysis methods were designed to handle (Pesnell et al., 2012; Stone et al., 1998; Fox et al., 2016). This data-rich environment has created both an opportunity and a bottleneck: while unprecedented information about the Sun–Earth system is available, extracting predictive knowledge from these heterogeneous, high-dimensional datasets poses a formidable challenge.

Artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools to address this challenge by learning complex, nonlinear relationships directly from data without requiring explicit analytical formulations. Over the past decade, AI methods—ranging from convolutional neural networks and recurrent architectures to ensemble learning and probabilistic models—have been increasingly applied to space weather problems such as solar flare prediction, CME detection and arrival time estimation, geomagnetic storm forecasting, and ionospheric disturbance modeling (Bobra & Couvidat, 2015; Wintoft et al., 2017; Chen et al., 2019; Sun et al., 2020). In many cases, AI-based approaches have demonstrated substantial improvements over climatological and empirical baselines, particularly in short-term forecasting and pattern recognition tasks (Camporeale, 2019).

However, the integration of AI into space weather science raises fundamental methodological questions. Purely data-driven models may lack physical interpretability, struggle with extrapolation beyond the training domain, and exhibit brittleness under rare extreme conditions—precisely when reliable forecasts are most critical (Camporeale et al., 2021; Liemohn et al., 2020). These concerns have motivated growing interest in hybrid AI–physics frameworks, physics-informed neural networks, and explainable AI techniques that aim to combine the predictive power of machine learning with the causal structure and conservation principles of space plasma physics (Karniadakis et al., 2021; Raissi et al., 2019). This review critically examines the current state of AI-driven space weather prediction, synthesizing advances across solar, heliospheric, and geospace domains. We evaluate benchmark performance against traditional physics-based approaches, assess the role of hybrid modeling and uncertainty quantification, and discuss emerging concepts such as space weather digital twins and autonomous forecasting systems. By situating AI within the broader epistemology of space weather science, this review aims to clarify both its transformative potential and its limitations, and to outline research directions necessary for robust, trustworthy, and operationally relevant forecasting.

This review is distinct from and substantially extends the authors' earlier work (Mishra et al., 2024), which examined cosmic ray flux variations and classical observational methods for space weather prediction using conventional data-analytic approaches. That study focused on empirical correlations between cosmic ray modulation and solar activity indicators, relying primarily on neutron monitor and satellite data interpreted through established physical frameworks. The present review, by contrast, addresses a fundamentally different scientific question: how artificial intelligence and machine learning architectures—including convolutional neural networks, recurrent and transformer models, physics-informed neural networks, and ensemble probabilistic methods—are reshaping the entire chain of space weather forecasting from solar event detection through geospace response. Rather than reviewing observational datasets or conventional prediction techniques, this work synthesizes AI benchmark performance across multiple forecasting tasks, critically evaluates hybrid AI–physics frameworks, addresses explainability and operational trust, and introduces the emerging paradigm of space weather digital twins. The scope, methodology, and intellectual contribution of this review are therefore qualitatively distinct from those of the 2024 publication. References to ground-based observatories and space weather dynamics in the current manuscript serve solely to provide necessary scientific context for evaluating AI-based methods, not to reproduce or repurpose prior descriptive material.

2 Problem Formulation

The effectiveness of artificial intelligence in space weather prediction is fundamentally constrained by the quality, diversity, and continuity of the underlying data ecosystem. Unlike many terrestrial AI applications, space weather modeling relies on heterogeneous observations spanning the solar interior, heliosphere, and geospace environment. These data streams differ markedly in spatial coverage, temporal cadence, physical meaning, and uncertainty characteristics, posing unique challenges for data-driven learning systems (Camporeale, 2019; Liemohn et al., 2020).

At the solar source region, AI models primarily exploit high-resolution imaging and magnetic field measurements from space-based observatories such as the Solar Dynamics Observatory, which provides continuous full-disk observations of the Sun in multiple extreme ultraviolet (EUV) and ultraviolet wavelengths, along with vector magnetograms from the Helioseismic and Magnetic Imager (HMI). These datasets enable machine learning models to characterize active region complexity, magnetic shear, and flux emergence—key precursors of solar flares and CMEs (Pesnell et al., 2012; Bobra & Couvidat, 2015). Complementary coronagraph observations from the Solar and Heliospheric Observatory and the Solar Terrestrial Relations Observatory (STEREO) provide information on CME morphology and kinematics, which are critical for downstream propagation and arrival-time prediction (Vourlidas et al., 2019).

In the heliosphere, in situ measurements of solar wind plasma and interplanetary magnetic field (IMF) parameters form the backbone of geomagnetic storm forecasting. Long-running missions such as the Advanced Composition Explorer and WIND have produced multi-decade time series of solar wind velocity, density, temperature, and IMF orientation at the L1 point, enabling the training of recurrent neural networks and transformer-based models for Dst and Kp prediction (Stone et al., 1998; Wintoft et al., 2017). More recent missions, including the Parker Solar Probe and Solar Orbiter, extend observations closer to the Sun, offering unprecedented opportunities for AI-based discovery of solar wind acceleration and CME initiation processes, albeit with challenges related to sparse temporal coverage and evolving measurement geometries (Fox et al., 2016; Müller et al., 2020).

The geospace domain introduces additional complexity through dense but unevenly distributed ground- and space-based observations. Global magnetometer networks provide continuous measurements of geomagnetic field perturbations, from which indices such as Dst, Kp, and AE are derived and widely used as AI training targets (Pulkkinen et al., 2017). Simultaneously, ionospheric observations from GNSS receiver networks yield total electron content (TEC) maps with high temporal resolution, supporting deep learning–based ionospheric forecasting and scintillation prediction (Sun et al., 2020; Deng et al., 2021). While these datasets are information-rich, they are also subject to regional biases, missing data, and instrumental inconsistencies that can degrade model robustness if not carefully addressed.

A defining challenge of the space weather data ecosystem is non-stationarity across solar cycles. Statistical properties of solar and geospace parameters vary significantly between solar minima and maxima, and extreme events are inherently rare. Consequently, AI models trained on one solar cycle may generalize poorly to another, a problem exacerbated by the limited number of truly extreme geomagnetic storms available for supervised learning (Owens & Forsyth, 2013; Camporeale et al., 2021). Addressing this issue requires advanced strategies such as data augmentation, transfer learning, self-supervised representation learning, and physics-informed constraints that reduce reliance on purely empirical correlations.

Equally important are issues of data provenance, latency, and interoperability in operational settings. Agencies such as the NOAA Space Weather Prediction Center depend on near-real-time data streams with strict latency requirements, meaning that AI models must be trained and validated on data products consistent with operational availability rather than idealized research archives (Schrijver et al., 2015). This operational constraint often leads to a trade-off between data richness and forecast timeliness, reinforcing the need for AI systems that are robust to incomplete and noisy inputs.

From a broader perspective, the space weather data ecosystem exemplifies the transition from data scarcity to data complexity. The key bottleneck is no longer the lack of observations, but the integration of multi-modal, multi-scale datasets into coherent learning frameworks. In my view, future progress will depend less on marginal increases in data volume and more on intelligent data curation, standardized benchmarks, and physically informed data fusion, enabling AI models to extract causal insight rather than spurious correlations. Solar imaging (EUV, X-ray, magnetograms)

3 AI Architectures and Forecasting Tasks

Artificial intelligence applications in space weather forecasting can be systematically understood by mapping AI architectures to specific physical prediction tasks across the Sun–Earth system. Unlike monolithic prediction problems, space weather forecasting is inherently modular, encompassing solar event detection, heliospheric propagation, and geospace response. Each task imposes distinct requirements on model architecture, input representation, temporal context, and uncertainty handling (Camporeale, 2019; Liemohn et al., 2020).

3.1 Convolutional Neural Networks for Solar Event Forecasting

Convolutional neural networks (CNNs) are the dominant architecture for image-based solar forecasting, particularly for solar flare and CME precursor detection. CNNs excel at extracting hierarchical spatial features from high-dimensional data such as vector magnetograms and extreme ultraviolet (EUV) images obtained from missions like the Solar Dynamics Observatory. By learning spatial correlations associated with magnetic complexity, flux emergence, and polarity inversion lines, CNNs enable automated flare classification and probability estimation (Bobra & Couvidat, 2015; Nishizuka et al., 2018).

The feature-extraction operation performed by each convolutional layer is expressed formally as:

$$h_{i,j}^{\ell} = \sigma\left(\sum_{p,q} W_{p,q}^{\ell} \cdot h_{i+p,j+q}^{\ell-1} + b^{\ell}\right) \quad (1)$$

where $h^{\ell}_{i,j}$ denotes the activation map at spatial position (i, j) in layer ℓ; $W^{\ell}_{p,q}$ is the learnable convolutional kernel at offset (p, q); b^{ℓ} is the layer bias; and σ(·) is a nonlinear activation such as ReLU. Applied to vector magnetogram inputs, successive layers build representations of increasingly complex magnetic features—from local flux gradients at shallow layers to polarity inversion line topology and δ-spot geometry at deeper layers—providing the physical basis for the high TSS values achieved by CNN-based flare classifiers.

Empirical benchmarks indicate that CNN-based flare predictors achieve true skill statistics (TSS) in the range of 0.6–0.8 for M- and X-class events with lead times of 24–48 hours, substantially outperforming climatological and persistence models (Leka et al., 2019). For CME detection, CNNs applied to coronagraph imagery from SOHO and STEREO have demonstrated reliable identification of CME onset and angular width, forming a critical input for downstream propagation models (Vourlidas et al., 2019). However, CNNs remain sensitive to training set imbalance and may overemphasize visually salient but physically secondary features unless carefully regularized.

3.2 Recurrent Neural Networks and LSTM Models for Geomagnetic Storm Prediction

Forecasting geomagnetic indices such as Dst and Kp requires models capable of learning temporal dependencies and delayed solar wind–magnetosphere coupling. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) architectures, are well suited to this task due to their ability to retain information over extended time horizons (Wintoft et al., 2017; Bala & Reiff, 2012).

The LSTM architecture resolves the vanishing-gradient problem of vanilla RNNs through three multiplicative gates. At each time step t, given input $x_t \in \mathbb{R}^d$ and the previous hidden state $h_{t-1} \in \mathbb{R}^h$, the cell computes:

$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big)$$
(2) forget gate

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
(3) input gate

$$C_t = tanh(W_C[h_{t-1}, x_t] + b_C)$$
(4) candidate cell

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t$$
(5) cell-state update

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
(6) output gate

$$h_t = o_t \odot tanh(C_t)$$
(7) hidden-state output

Here W_f, W_i, W^A, $W_o \in \mathbb{R}^{h\times(d+h)}$ are learned weight matrices; b denotes the corresponding bias vectors; σ is the sigmoid function; and ⊙ is the Hadamard (element-wise) product. The cell state C_t acts as an explicit memory register whose content is selectively preserved by the forget gate f_t and updated by the input gate i_t—a structure that directly mirrors the ring-current energy balance described by the classical Burton et al. (1975) coupling ODE:

$$\frac{dDst*}{dt} = Q(E_{sw}) - \frac{Dst*}{\tau} \quad (8)$$

where p* denotes the pressure-corrected Dst index, $Q(E_{sw}) = a_0 E_{sw} + b_0$ is the solar-wind energy injection function of the dawn-dusk electric field $E\square_v = v\square_v B\square$ (for southward IMF), and $\tau \approx 7.7$ h is the ring-current decay time. The LSTM learn a fully nonlinear, history-dependent generalization of $Q(\cdot)$, capturing delayed energy injection and asymmetric storm recovery that the linear Burton form cannot represent.

LSTM-based models trained on upstream solar wind parameters (velocity, density, IMF Bz) measured at L1 by missions such as ACE and WIND consistently achieve correlation coefficients exceeding 0.9 for 1-hour Dst forecasts, with useful skill extending to 6–12 hours (Temerin & Li, 2006; McGranaghan et al., 2017). These models effectively learn nonlinear energy transfer processes that are difficult to encode explicitly in empirical coupling functions. Nevertheless, their performance degrades during extreme geomagnetic storms, reflecting limited representation of rare events in historical datasets.

3.3 Transformer Architectures for Multivariate Time Series Forecasting

More recently, transformer-based architectures have gained attention in space weather research due to their ability to model long-range temporal dependencies using self-attention mechanisms. Unlike RNNs, transformers process entire time windows in parallel, improving training efficiency and enabling scalable multivariate forecasting (Vaswani et al., 2017; Willard et al., 2022).

The mathematical core of the transformer is the scaled dot-product self-attention operation. Given an input sequence of length T projected into query matrix $Q \in \mathbb{R}^{T\times dk}$, key matrix $K \in \mathbb{R}^{T\times dk}$, and value matrix $V \in \mathbb{R}^{T\times d\beta}$, the attention output is:

$$(Q,K,V) = \left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{9}$$

The $\sqrt{d\square}$ scaling factor prevents the dot products from growing large enough to push the softmax into a flat gradient region. Multi-head attention runs h parallel attention layers over distinct learned linear projections of Q, K, V, then concatenates their outputs:

$$(Q,K,V) = W^{O}[head_1, \dots, head_h] \tag{10}$$

with $head_i = \text{Attention}(QW^{k}_{i}, KW^{I}_{i}, VW^{\beta}_{i})$ and $W^{k}_{i} \in \mathbb{R}^{dod_i emh}$, $W^{I}_{i} \in \mathbb{R}^{dod_i emh}$, $W^{\beta}_{i} \in \mathbb{R}^{dod_i emh}$, $W^{N} \in \mathbb{R}^{hdo}$ as projection matrices. In space weather applications each attention head can specialize in a distinct temporal pattern of the solar-wind time series—one head attending to sudden southward IMF turnings, another to prolonged enhanced solar wind pressure—so that the multi-head output integrates all geomagnetically relevant features across multi-day windows simultaneously.

Initial studies applying transformers to geomagnetic index prediction and ionospheric TEC forecasting report improved stability and reduced error accumulation at longer lead times compared to LSTM baselines (Sun et al., 2020; Deng et al., 2021). Their capacity to integrate heterogeneous inputs—solar wind data, geomagnetic indices, and seasonal or solar-cycle indicators—makes transformers particularly attractive for next-generation operational systems. However, their large parameter counts increase the risk of overfitting and necessitate careful validation under out-of-distribution conditions.

3.4 Ensemble Learning and Probabilistic Forecasting

Operational space weather forecasting demands not only accuracy but also reliable uncertainty quantification. Ensemble learning methods—such as random forests, gradient boosting, and multi-model neural ensembles have been widely adopted to produce probabilistic forecasts and reduce model variance (Chen et al., 2019; Crown, 2012). These approaches are especially valuable for CME arrival time prediction, where ensembles can capture sensitivity to uncertain initial conditions.

Probabilistic outputs are increasingly favored by agencies such as the NOAA Space Weather Prediction Center, as they support risk-based decision-making for satellite operators and power grid managers (Pulkkinen et al., 2017). Bayesian neural networks and Gaussian process regression further extend this paradigm by explicitly modeling predictive uncertainty, though their computational cost can limit real-time deployment.

3.5 Physics-Informed and Hybrid AI Architectures

A key limitation of purely data-driven AI models is their weak physical interpretability and limited extrapolation capability. To address this, physics-informed neural networks (PINNs) and hybrid AI–physics architectures incorporate governing equations, conservation laws, or physically motivated constraints directly into the learning process (Raissi et al., 2019; Karniadakis et al., 2021).

To make the hybrid framework mathematically concrete, we first state the ideal MHD equation system that operational solvers (ENLIL, BATS-R-US, OpenGGCM) integrate. For plasma mass density ρ, bulk velocity v, magnetic field B, thermal pressure p, and total energy density e, the conservation laws are:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0$$

(11) continuity

$$\frac{\partial (\rho v)}{\partial t} + \nabla \cdot (\rho v v^{T}) + \nabla p* - (B \cdot \nabla)B = 0$$

(12) momentum [$p^* = p + B^2/2\mu_0$]

$$\frac{\partial B}{\partial t} - \nabla \times (v \times B) = 0, \quad \nabla \cdot B = 0$$

(13) induction + solenoidal constraint

$$\frac{\partial e}{\partial t} + \nabla \cdot [(e + p*)v - (v \cdot B)B] = 0$$

(14) energy [$e = \rho v^2/2 + p/(\gamma-1) + B^2/2\mu_0$]

Equations (11)–(14) govern the evolution of coronal and solar-wind plasma from the inner boundary (~0.1 AU) through the heliosphere. Uncertainty in boundary conditions at this inner boundary is the dominant source of CME arrival-time forecast error, and it is precisely this uncertainty that hybrid AI systems reduce by learning physically plausible boundary state estimates from observational proxies.

In the PINN framework (Raissi et al., 2019), a neural network û(x; θ) with parameters θ is trained to minimise a composite loss that enforces both observational fidelity and PDE residual:

$$\mathcal{L}(\theta) = \mathcal{L}_{data} + \lambda \mathcal{L}_{phys}$$

(15)

$$\mathcal{L}_{data} = \frac{1}{N}\sum_{i=1}^{N} \| \hat{u}_i - u_i \|^2$$

(16) data loss

$$\mathcal{L}_{phys} = \frac{1}{N_r}\sum_{i=1}^{N_r} \| \wp[\hat{u}_i] \|^2$$

(17) physics loss

Here $\mathcal{L}_{DaTa}$ measures mean-squared error against N labeled in situ observations; $\mathcal{L}_{phys}$ sums the squared PDE residual $\wp[\hat{u}]$—the left-hand side of Equations (11)–(14) evaluated on the predicted field—at N_r collocation points sampled from the domain; and λ > 0 balances the two objectives. By minimising $\mathcal{L}(\theta)$ jointly the PINN learns a plasma-state field that is simultaneously consistent with sparse measurements and the conservation laws of MHD, enabling physically plausible interpolation in the poorly observed inner heliosphere.

In space weather applications, hybrid models typically use AI to estimate uncertain inputs—such as CME speed, orientation, or solar wind boundary conditions—which are then ingested by magnetohydrodynamic (MHD) solvers. This division of labor leverages the pattern-recognition strength of AI while preserving the causal structure of physical models (Riley et al., 2018; Camporeale et al., 2021). In my view, such hybridization represents the most scientifically robust pathway for operational forecasting, particularly for extreme and unprecedented events.

3.6 Mapping Architectures to Forecasting Tasks

The diversity of AI architectures reflects the diversity of space weather forecasting tasks. CNNs dominate solar image analysis, RNNs and transformers excel in temporal geospace prediction, and ensemble and probabilistic models underpin operational decision support. No single architecture is universally optimal; instead, forecasting skill emerges from task-specific model selection and intelligent integration across the Sun–Earth chain.

4 Benchmark Comparisons

Benchmarking is essential for assessing whether artificial intelligence provides measurable, reproducible improvements over traditional physics-based and empirical space weather models. Unlike many machine learning domains, space weather benchmarking is complicated by non-stationary data distributions, event rarity, heterogeneous evaluation metrics, and inconsistent test periods across studies. Consequently, meaningful comparisons require careful alignment of datasets, lead times, and performance metrics (Liemohn et al., 2020; Camporeale, 2019).

4.1 Benchmarking Philosophy and Evaluation Metrics

Benchmark studies in space weather typically evaluate models using metrics tailored to specific forecasting tasks:

- Solar flare prediction: True Skill Statistic (TSS), Heidke Skill Score (HSS), Probability of Detection (POD), False Alarm Rate (FAR)
- CME arrival time: Mean Absolute Error (MAE), Root Mean Square Error (RMSE)
- Geomagnetic indices (Dst, Kp): Correlation coefficient (CC), RMSE
- Ionospheric TEC forecasting: Percentage error, RMSE

Crucially, AI models are often compared against climatology, persistence, empirical coupling functions, and full MHD models, rather than against each other alone (Schrijver et al., 2015; Pulkkinen et al., 2017).

4.2 Solar Flare Forecasting Benchmarks

Solar flare prediction has emerged as one of the most mature benchmarking areas for AI in space weather. CNN-based models trained on vector magnetograms from the Solar Dynamics Observatory routinely outperform traditional statistical methods.

Across multiple studies, AI-based flare predictors achieve TSS values between 0.6 and 0.8 for M- and X-class flares with 24–48 hour lead times, compared with 0.3–0.45 for climatological and rule-based methods (Bobra & Couvidat, 2015; Nishizuka et al., 2018; Leka et al., 2019). Importantly, benchmark results indicate that performance gains saturate unless class imbalance and temporal leakage are explicitly addressed, highlighting the importance of rigorous validation protocols.

4.3 CME Arrival Time Prediction Benchmarks

Forecasting CME arrival time remains a central operational challenge. Physics-based heliospheric models such as ENLIL typically yield arrival-time errors of 18–24 hours, largely due to uncertainty in CME initial speed, direction, and magnetic structure (Riley et al., 2018; Manchester et al., 2017). AI and ensemble ML models that estimate CME kinematics directly from coronagraph imagery reduce MAE to 8–12 hours, representing a 40–60% improvement over physics-only baselines (Chen et al., 2019; Wiener et al., 2020). Hybrid AI–physics systems, in which AI-derived CME parameters initialize MHD models, consistently outperform both standalone approaches, underscoring the value of model coupling.

4.4 Geomagnetic Storm (Dst, Kp) Forecast Benchmarks

Geomagnetic storm prediction is a key success area for time-series AI models. LSTM and transformer-based architectures trained on upstream solar wind measurements from ACE and WIND achieve correlation coefficients exceeding 0.9 for 1-hour-ahead Dst forecasts, compared to 0.7–0.8 for empirical coupling functions such as Burton-type models (Temerin & Li, 2006; Wintoft et al., 2017). For Kp forecasting, ensemble ML models reduce RMSE from 1.2–1.5 (physics/empirical models) to 0.6–0.9, while also lowering false alarm rates during quiet geomagnetic conditions (McGranaghan et al., 2017; Crown, 2012). However, benchmark analyses show that AI performance degrades disproportionately during extreme storms (Dst < −250 nT), reflecting sparse training data for rare events (Pulkkinen et al., 2017).

4.5 Ionospheric TEC Forecasting Benchmarks

Deep learning models applied to ionospheric TEC prediction demonstrate **20–40% error reduction** relative to climatological and physics-based ionospheric models, particularly at regional scales and short lead times (Sun et al., 2020; Deng et al., 2021). Transformer-based models outperform LSTM baselines in multi-hour forecasting, suggesting superior handling of long-range temporal dependencies. Nevertheless, benchmarks also reveal strong sensitivity to data density, with degraded performance in regions lacking dense GNSS receiver coverage—an important limitation for global operational deployment.

4.6 Summary of Quantitative Benchmarks

Table 1. Representative Benchmark Comparisons of Space Weather Forecasting Models

Forecast Task	**Metric**	**Physics / Empirical Models**	**AI-Based Models**	**Hybrid AI–Physics**
Solar flare (M/X)	TSS	0.30–0.45	**0.60–0.80**	0.65–0.85
CME arrival time	MAE (h)	18–24	**8–12**	**6–10**
Dst (1 h ahead)	CC	0.70–0.80	**0.90–0.95**	0.92–0.96
Kp index	RMSE	1.2–1.5	**0.6–0.9**	0.5–0.8
TEC forecasting	% error	25–35%	**15–20%**	12–18%

4.7 Critical Assessment and Limitations of Benchmarks

While benchmark results clearly demonstrate the added value of AI, several caveats must be emphasized. Many published benchmarks rely on retrospective datasets with limited inter-solar-cycle validation, potentially inflating performance estimates (Liemohn et al., 2020). Differences in event definitions, lead times, and test periods further complicate cross-study comparison. In my view, the field urgently needs standardized, community-endorsed benchmark datasets and evaluation protocols, analogous to those used in climate modeling and computer vision.

5 Hybrid AI–Physics Frameworks

Hybrid AI–physics frameworks have emerged as a dominant paradigm for advancing space weather prediction, addressing fundamental limitations of purely data-driven or purely physics-based approaches. The core motivation is epistemic: while first-principles magnetohydrodynamic (MHD) models encode causal structure and conservation laws, they are hampered by uncertain initial and boundary conditions, parameterizations, and computational expense. Conversely, AI models excel at pattern recognition and rapid inference but may violate physical constraints and extrapolate poorly during extreme or unprecedented events. Hybrid frameworks aim to combine the generalization and interpretability of physics with the adaptivity and efficiency of AI (Camporeale, 2019; Liemohn et al., 2020).

5.1 AI-Assisted Parameter Estimation for Physics Models

One of the most mature hybrid strategies uses AI to infer latent or poorly observed parameters that initialize or constrain physics-based models. In CME forecasting, for example, global heliospheric MHD models such as ENLIL require estimates of CME speed, direction, angular width, and magnetic orientation—parameters that are only indirectly observable from coronagraph imagery. Machine learning models trained on historical events can infer these quantities from multi-viewpoint observations, significantly reducing downstream forecast error (Riley et al., 2018; Chen et al., 2019).

Benchmark studies demonstrate that AI-initialized MHD simulations reduce CME arrival-time mean absolute error from 18–24 hours to approximately 6–10 hours, outperforming both standalone AI regressors and physics-only simulations (Manchester et al., 2017; Wiener et al., 2020). This division of labor—AI for inference, physics for propagation—has proven particularly effective in operational contexts.

5.2 Physics-Informed Neural Networks (PINNs)

Physics-informed neural networks (PINNs) represent a more tightly coupled hybrid approach, embedding governing equations directly into the loss function of neural networks. By penalizing violations of conservation laws (e.g., mass, momentum, energy, and magnetic flux), PINNs constrain the hypothesis space of the model and improve physical consistency (Raissi et al., 2019; Karniadakis et al., 2021).

In space weather applications, PINNs have been explored for reduced-order modeling of plasma dynamics, data assimilation in sparsely observed regions, and surrogate modeling of computationally expensive MHD solvers. While still largely experimental, PINNs offer a promising route to generalizable learning under limited data, particularly for extreme geomagnetic storms where historical examples are scarce (Camporeale et al., 2021).

5.3 AI-Based Emulation of Physics Solvers

Another hybrid strategy involves using AI as a surrogate or emulator for selected components of physics-based models. Deep neural networks can approximate the input–output behavior of MHD or ionospheric models at a fraction of the computational cost, enabling rapid ensemble forecasting and uncertainty quantification (McGranaghan et al., 2017; Willard et al., 2022).

Such emulators are especially valuable for operational forecasting centers, where computational latency constrains ensemble size. However, their reliability depends critically on the representativeness of the training data and the stability of the underlying physical regime, necessitating careful validation and fallback mechanisms to full-physics solvers.

5.4 Coupled AI–Physics Data Assimilation

Hybrid frameworks also extend to data assimilation, where AI augments traditional Kalman filtering or variational methods by learning observation operators or error covariances. In the Sun–Earth system, where observations are sparse and unevenly distributed, AI-enhanced assimilation improves state estimation in the heliosphere and magnetosphere, directly benefiting short-term forecasting of geomagnetic disturbances (Lugaz et al., 2018; Liemohn et al., 2020). From an operational standpoint, such systems are being explored by agencies including NASA and NOAA Space Weather Prediction Center as precursors to fully integrated space weather digital twins.

5.5 Strengths, Limitations, and Scientific Implications

Hybrid AI–physics frameworks consistently outperform standalone approaches across major benchmarks, particularly for CME arrival time and geomagnetic storm intensity. Their principal strengths include improved extrapolation, reduced unphysical behavior, and enhanced interpretability. However, they also introduce new layers of complexity, including model coupling uncertainty, error propagation across components, and challenges in attribution of forecast skill.

In my view, the long-term scientific significance of hybrid frameworks lies not merely in improved accuracy, but in their potential to accelerate physical discovery. By highlighting systematic discrepancies between learned representations and governing equations, hybrid models can expose missing physics and guide future theoretical development.

6 Explainability and Trust in Operational Settings

The transition of artificial intelligence from research prototypes to operational space weather forecasting hinges critically on explainability, reliability, and trust. Operational centers must justify forecasts that can trigger costly mitigation actions—such as satellite safe-mode transitions, power grid load shedding, or aviation rerouting—often under time pressure and with incomplete information. Consequently, black-box predictions without transparent rationale are insufficient for mission-critical decision-making (Schrijver et al., 2015; Pulkkinen et al., 2017).

6.1 Trust Requirements in Operational Space Weather

Operational trust is multi-dimensional. Forecasting agencies require (i) interpretability—clear explanations of why a forecast was issued; (ii) consistency—stable performance across solar conditions; (iii) uncertainty awareness—probabilistic outputs with calibrated confidence; and (iv) auditability—traceability of data sources, model versions, and decision logic (Liemohn et al., 2020). These requirements are particularly salient for agencies such as the NOAA Space Weather Prediction Center, where AI outputs are increasingly used alongside physics-based guidance in human-in-the-loop workflows.

6.2 Explainable AI (XAI) Techniques for Space Weather

Explainable AI (XAI) methods aim to reveal how AI models map inputs to predictions, enabling forecasters to assess physical plausibility. In solar flare prediction, saliency maps and layer-wise relevance propagation (LRP) applied to convolutional neural networks identify magnetogram regions that most strongly influence flare probability estimates. These highlighted regions frequently correspond to known physical drivers—such as strong magnetic gradients and polarity inversion lines—providing post hoc validation of learned representations (Bobra & Couvidat, 2015; Leka et al., 2019).

Feature-attribution methods such as SHAP (SHapley Additive exPlanations) and permutation importance are widely used in time-series models for geomagnetic storm prediction. These techniques consistently rank southward interplanetary magnetic field (IMF Bz), solar wind speed, and dynamic pressure among the most influential predictors of Dst and Kp, in agreement with decades of magnetospheric physics (Temerin & Li, 2006; McGranaghan et al., 2017). Such concordance between AI explanations and physical understanding is a key enabler of trust.

6.3 Uncertainty Quantification and Probabilistic Forecasts

Explainability alone is insufficient without explicit uncertainty quantification. Operational users must know not only what is predicted, but how confident the model is. Ensemble learning, Bayesian neural networks, and Gaussian process regression provide probabilistic forecasts that express uncertainty arising from data noise, model structure, and input variability (Chen et al., 2019; Camporeale et al., 2021).

In operational trials, probabilistic AI forecasts have been shown to reduce false alarms during quiet periods while maintaining sensitivity to severe events—an essential balance for infrastructure protection (Pulkkinen et al., 2017). However, improper calibration can lead to overconfident predictions, underscoring the need for continuous verification and reliability diagrams tailored to space weather use cases.

6.4 Human–AI Collaboration in Forecast Operations

Rather than replacing human forecasters, AI systems are increasingly deployed as decision-support tools within human–AI teams. Forecasters combine AI guidance with physical intuition, contextual knowledge, and cross-model comparison to produce final advisories. Studies indicate that trust is maximized when AI systems provide interpretable cues and uncertainty ranges, enabling experts to override or contextualize model outputs when necessary (Schrijver et al., 2015; Liemohn et al., 2020). The most robust operational paradigm is human-centered AI, where explainability is not an afterthought but a design requirement. Systems optimized solely for statistical accuracy risk eroding trust if their behavior cannot be reconciled with physical reasoning or operational experience.

6.5 Governance, Validation, and Ethical Considerations

Explainability also underpins governance and accountability. Operational AI systems must support retrospective analysis following missed events or false alarms, enabling identification of failure modes and model biases. This is particularly important given the societal and economic consequences of space weather forecasts, which can involve cross-border infrastructure and international coordination (Eastwood et al., 2017). Standardized validation protocols, open benchmark datasets, and transparent reporting of model assumptions are therefore essential. Without such measures, AI adoption risks fragmenting operational practices and undermining confidence among stakeholders. Explainability and trust are not auxiliary concerns but core scientific and operational challenges in AI-based space weather forecasting. The convergence of XAI, probabilistic modeling, and human-in-the-loop design provides a viable pathway toward trustworthy AI systems. Ultimately, the success of AI in operational settings will be judged not by marginal gains in forecast accuracy alone, but by its ability to earn and sustain the confidence of forecasters, decision-makers, and the societies they serve.

7 Toward Space Weather Digital Twins

The concept of a space weather digital twin represents a natural evolution of hybrid AI–physics frameworks toward an integrated, continuously updating representation of the Sun–Earth system. Borrowed from aerospace and manufacturing engineering, a digital twin is a high-fidelity virtual counterpart of a physical system that assimilates real-time data, executes predictive simulations, and supports decision-making under uncertainty (Glaessgen & Stargel, 2012; Tuegel et al., 2011). In the context of space weather, a digital twin aims to couple solar, heliospheric, and geospace models into a unified, AI-augmented forecasting ecosystem capable of near-real-time situational awareness and anticipatory prediction.

7.1 Conceptual Architecture of a Space Weather Digital Twin

A space weather digital twin integrates four core components: (i) continuous data ingestion, (ii) physics-based simulation, (iii) AI-based emulation and inference, and (iv) decision-support interfaces. Observational data streams from solar observatories, heliospheric probes, and ground-based sensor networks are assimilated in near real time to update the system state. Physics-based models—typically magnetohydrodynamic (MHD) solvers for the heliosphere and coupled magnetosphere-ionosphere models for geospace—provide causal structure and physically consistent evolution (Tóth et al., 2005; Manchester et al., 2017). Artificial intelligence plays a dual role within this architecture. First, AI accelerates computation by emulating expensive model components, enabling ensemble forecasting and uncertainty quantification under operational time constraints. Second, AI performs inference tasks—such as estimating CME parameters or identifying precursors of geomagnetic storms—that are difficult to constrain directly from observations (Camporeale et al., 2021; McGranaghan et al., 2022). The resulting system is not static but adaptive, continuously recalibrated as new observations arrive.

7.2 Data Assimilation and Closed-Loop Forecasting

A defining feature of digital twins is closed-loop operation, in which forecast errors are systematically fed back into the system to improve subsequent predictions. In space weather applications, this involves AI-enhanced data assimilation schemes that learn observation operators and error covariances for sparse and heterogeneous datasets (Liemohn et al., 2020). Unlike traditional open-loop forecasting, closed-loop digital twins can self-correct, adjusting model states and parameters as discrepancies between predictions and observations emerge. Such closed-loop capability is particularly valuable during extreme events, when rapid system evolution and nonlinear feedbacks challenge conventional forecasting approaches. By continuously integrating upstream solar wind data and geospace observations, a digital twin can update storm intensity forecasts and impact assessments in near real time, extending actionable lead time for mitigation measures.

7.3 Autonomous and Scenario-Based Forecasting

Beyond real-time prediction, space weather digital twins enable counterfactual and scenario-based analysis. AI-assisted twins can simulate hypothetical solar eruptions, altered CME trajectories, or worst-case geomagnetic conditions to assess infrastructure vulnerability and resilience. This capability is critical for stress-testing satellite constellations, power grids, and navigation systems against low-probability, high-impact events such as Carrington-class storms (Baker et al., 2018; Oughton et al., 2017). From an operational perspective, such scenario exploration supports a shift from reactive forecasting toward anticipatory risk management, aligning space weather services with broader disaster preparedness frameworks.

7.4 Institutional and Operational Context

Major space agencies and operational centers, including NASA, the European Space Agency, and the NOAA Space Weather Prediction Center, have identified digital twins as a strategic priority for next-generation space weather services. These efforts emphasize interoperability, modular design, and transparency to ensure that digital twins can integrate new models, sensors, and AI components as capabilities evolve. However, realizing a fully operational space weather digital twin remains a formidable challenge. It requires sustained investment in data infrastructure, standardized interfaces, rigorous validation, and governance frameworks that define responsibility and accountability for AI-assisted decisions.

7.5 Scientific and Epistemic Implications

From a scientific standpoint, space weather digital twins blur the traditional boundary between model and observation. By embedding AI within the modeling loop, digital twins act as active scientific instruments, not merely predictive tools. Persistent mismatches between twin predictions and observations can reveal gaps in physical understanding, guiding targeted observational campaigns and theoretical development. The most profound contribution of space weather digital twins will be epistemic rather than purely operational. They offer a framework for unifying data, theory, and computation into a living model of the Sun–Earth system—one that evolves with both the physical environment and our scientific understanding of it.

8 Societal and Economic Implications

Space weather is no longer a niche concern of heliophysics; it represents a systemic socio-technical risk to modern civilization. The increasing reliance on space-based assets, electrified infrastructure, and high-precision navigation has amplified societal exposure to solar-driven disturbances. As a result, improvements in space weather forecasting—particularly through AI-enhanced systems—carry significant economic, safety, and governance implications (Schrijver et al., 2015; Eastwood et al., 2017).

8.1 Economic Costs of Space Weather Hazards

Quantitative assessments indicate that severe space weather events can impose extraordinary economic losses. A Carrington-class geomagnetic storm is estimated to cause USD 1–2 trillion in global damages within the first year, with recovery times extending over multiple years due to transformer damage and supply-chain disruptions (Oughton et al., 2017; Baker et al., 2018). Even moderate geomagnetic storms generate recurring costs through satellite anomalies, communication outages, and degraded navigation accuracy. In the satellite sector alone, space weather is a leading cause of on-orbit anomalies. Surface charging, deep dielectric charging, and single-event effects driven by energetic particles can shorten satellite lifetimes or trigger permanent failures, translating into hundreds of millions of dollars in annual losses globally (Schrijver et al., 2015). AI-enabled early warning systems that provide probabilistic risk estimates can reduce these losses by enabling timely operational mitigations, such as spacecraft reorientation or temporary shutdown of sensitive components.

8.2 Critical Infrastructure and Energy Systems

Electric power grids are particularly vulnerable to geomagnetically induced currents (GICs), which can saturate transformers, cause voltage instability, and lead to cascading failures. The 1989 Quebec blackout remains a canonical example of how space weather can disrupt large-scale energy infrastructure within minutes (Pulkkinen et al., 2017). As grids become more interconnected and heavily loaded, their susceptibility to space weather–driven disturbances increases. AI-enhanced space weather forecasts support grid resilience by extending actionable lead times and improving storm intensity estimates. From a socio-economic perspective, even modest reductions in false alarms and missed events yield disproportionately large economic benefits, as unnecessary grid shutdowns and equipment stress carry significant costs (Oughton et al., 2017). In this context, hybrid AI–physics models and digital twins serve as decision-support tools rather than deterministic predictors.

8.3 Aviation, Navigation, and Human Safety

Global navigation satellite systems (GNSS) underpin aviation navigation, maritime operations, precision agriculture, and financial time-stamping. Space weather–induced ionospheric disturbances degrade GNSS accuracy and availability, particularly at high latitudes, increasing operational risk for aviation and polar routes (Deng et al., 2021). Solar energetic particle events further pose radiation hazards to aircrew and passengers on long-haul polar flights. AI-driven ionospheric forecasting and radiation nowcasting improve situational awareness and enable dynamic route planning, reducing both safety risks and fuel costs. While the economic impact of individual events may appear modest, their cumulative effect across global aviation operations is substantial, reinforcing the value of reliable AI-assisted space weather services.

8.4 Societal Resilience and Risk Governance

Beyond direct economic losses, space weather events have societal resilience implications. Prolonged power outages, communication failures, and navigation disruptions can impair emergency response, healthcare delivery, and financial systems. These cascading effects highlight the need to treat space weather as a systemic risk, comparable to earthquakes or extreme weather events, within national and international resilience planning (Eastwood et al., 2017). AI-enabled space weather digital twins provide a mechanism for stress-testing infrastructure and evaluating preparedness strategies under hypothetical extreme scenarios. Such scenario-based analysis supports evidence-based policy-making and investment decisions, enabling governments and utilities to prioritize mitigation measures where they deliver the greatest societal benefit.

8.5 Equity, Responsibility, and Ethical Considerations

The societal benefits of improved space weather prediction are unevenly distributed. Regions with sparse observational coverage or limited infrastructure investment may face disproportionate impacts from space weather hazards while benefiting least from advanced forecasting capabilities. This raises ethical questions regarding data sharing, international cooperation, and equitable access to space weather services (Schrijver et al., 2015). From my perspective, AI should be viewed as a public-good technology in this domain. Open data policies, transparent models, and internationally coordinated forecasting frameworks are essential to ensure that AI-driven advances in space weather prediction enhance global resilience rather than exacerbate existing inequalities.

The societal and economic implications of space weather underscore the importance of transitioning from reactive forecasting to anticipatory, risk-informed decision support. AI does not merely improve forecast accuracy; it reshapes how space weather risk is understood, communicated, and managed across sectors. When embedded within trustworthy, explainable, and hybrid AI–physics systems, AI-enhanced space weather prediction offers tangible benefits for economic stability, public safety, and long-term societal resilience.

9 Limitations and Open Challenges

Despite rapid progress, the application of artificial intelligence to space weather prediction faces fundamental scientific, technical, and operational limitations. These challenges constrain both forecast reliability and the pace at which AI-driven methods can be adopted in mission-critical environments. Understanding these limitations is essential not only for responsible deployment, but also for identifying priority research directions (Camporeale, 2019; Liemohn et al., 2020).

9.1 Data Sparsity and Class Imbalance

A persistent limitation of AI-based space weather forecasting is the scarcity of extreme events in historical records. Carrington-class geomagnetic storms, major solar energetic particle events, and fast Earth-directed CMEs occur too infrequently to support conventional supervised learning approaches. As a result, AI models tend to be optimized for moderate conditions and may systematically underperform during the most consequential events (Pulkkinen et al., 2017; Baker et al., 2018). Class imbalance exacerbates this issue. For example, M- and X-class solar flares represent a small fraction of all solar activity, leading to biased classifiers unless reweighting, resampling, or cost-sensitive learning strategies are applied (Leka et al., 2019). Even with mitigation, performance metrics can appear deceptively high while masking poor recall for rare but high-impact events.

9.2 Non-Stationarity across Solar Cycles

The Sun–Earth system is non-stationary on multiple time scales, particularly across the ~11-year solar cycle. Statistical properties of solar magnetic fields, solar wind conditions, and geomagnetic responses vary substantially between solar minima and maxima. AI models trained on data from one solar cycle often exhibit degraded performance when applied to another, limiting their long-term generalizability (Owens & Forsyth, 2013; Camporeale et al., 2021). This challenge is compounded by instrumental evolution and mission turnover, which introduce systematic shifts in data characteristics over time. Addressing non-stationarity requires adaptive learning strategies, transfer learning, and continual model updating—capabilities that are not yet standard in operational space weather systems.

9.3 Limited Observability and Sparse Spatial Coverage

Unlike terrestrial weather systems, space weather observations are sparse, unevenly distributed, and often indirect. Key regions of the heliosphere and magnetosphere remain poorly instrumented, forcing AI models to infer system state from limited vantage points. This limited observability constrains forecast skill and increases uncertainty, particularly for CME magnetic structure and geoeffectiveness (Manchester et al., 2017; Vourlidas et al., 2019). While AI can partially compensate through learned correlations, it cannot fully replace missing physical information. This underscores the continued importance of targeted observational missions and sensor network design, potentially guided by AI-based observing system simulation experiments.

9.4 Model Robustness and Out-of-Distribution Behavior

AI models are notoriously vulnerable to out-of-distribution (OOD) inputs, producing confident but incorrect predictions when confronted with conditions not represented in the training data. In space weather, such conditions may arise during extreme solar activity, sensor degradation, or data outages. Ensuring robust behavior under these circumstances remains an open challenge (Liemohn et al., 2020). Current approaches such as uncertainty thresholds, ensemble disagreement, and fallback to physics-based models mitigate but do not eliminate this risk. Robustness must therefore be treated as a first-class design objective rather than a post hoc correction.

9.5 Interpretability and Scientific Insight

Although explainable AI techniques have improved transparency, deep learning models still struggle to provide causal explanations comparable to physical theory. Feature attribution methods can identify influential inputs but do not necessarily reveal mechanistic understanding. This limits the ability of AI models to contribute directly to scientific discovery and complicates their acceptance within the heliophysics community (Camporeale, 2019). Bridging this gap requires closer integration between AI development and theoretical modeling, with a focus on learning physically meaningful representations rather than maximizing short-term predictive performance.

9.6 Validation, Reproducibility, and Benchmarking Gaps

A further challenge lies in the lack of standardized benchmarks and validation protocols. Studies often use different datasets, time periods, and evaluation metrics, making cross-comparison difficult and sometimes misleading. Reproducibility is further hindered by proprietary data, undocumented preprocessing steps, and rapidly evolving software environments (Liemohn et al., 2020; Schrijver et al., 2015). Establishing community-agreed benchmark datasets and open evaluation frameworks is essential for translating research advances into operational trust.

9.7 Governance and Over-Reliance Risks

Finally, the increasing automation enabled by AI raises governance concerns. Over-reliance on algorithmic forecasts—particularly opaque or poorly validated systems—could amplify risk rather than reduce it, especially if human oversight is diminished. Clear accountability, transparent decision-making processes, and defined roles for human forecasters remain indispensable (Eastwood et al., 2017).

9.8 Synthesis of Open Challenges

Taken together, these limitations highlight that AI is not a panacea for space weather prediction. Its strengths must be balanced against inherent constraints arising from data scarcity, system complexity, and operational risk. Progress will depend on sustained investment in observations, hybrid AI–physics methods, standardized evaluation, and human-centered system design. Addressing these challenges is essential if AI is to fulfill its promise as a transformative tool for space weather science and societal resilience.

10 Future Directions

The next phase of research in AI-enabled space weather prediction will be defined not by incremental performance gains, but by conceptual advances in learning paradigms, system integration, and scientific governance. As space weather forecasting moves toward operational digital twins and autonomous decision-support systems, several interlinked research directions emerge as particularly critical.

10.1 Foundation Models for the Sun–Earth System

A promising frontier is the development of foundation models—large-scale, pre-trained AI models capable of learning universal representations from vast, heterogeneous heliophysical datasets. Analogous to foundation models in climate science and natural language processing, such systems would be trained on multi-decadal observations spanning solar imagery, heliospheric plasma measurements, geomagnetic indices, and ionospheric data (Reichstein et al., 2019; Willard et al., 2022). Foundation models could enable transfer learning across forecasting tasks, allowing knowledge gained from abundant moderate events to inform prediction of rare extremes. In my view, this paradigm is especially well suited to space weather, where data diversity is high but labeled extreme events are scarce. However, the computational cost and carbon footprint of training such models must be carefully balanced against scientific benefit.

10.2 Learning from Rare and Extreme Events

Addressing the prediction of low-probability, high-impact events remain one of the most pressing challenges. Future research must move beyond conventional supervised learning toward self-supervised, contrastive, and generative approaches that exploit unlabeled data and synthetic simulations (Camporeale et al., 2021). Physics-based simulations of extreme scenarios, combined with AI-driven data augmentation, offer a pathway to expanding the effective training space without overfitting to historical records. Generative models may also play a role in stress-testing infrastructure and forecasting systems under hypothetical Carrington-class conditions, supporting resilience planning even in the absence of direct observational analogs (Baker et al., 2018).

10.3 Adaptive and Continual Learning across Solar Cycles

Given the strong non-stationarity of the Sun–Earth system, continual learning frameworks that adapt across solar cycles are essential. Rather than retraining models episodically, future AI systems should update dynamically as new observations arrive, while avoiding catastrophic forgetting of past knowledge (Owens & Forsyth, 2013; Liemohn et al., 2020). This capability will require robust mechanisms for detecting distributional shifts, validating model updates in near real time, and maintaining operational stability. Continual learning also raises governance questions regarding version control, traceability, and accountability in operational forecasting.

10.4 AI-Guided Observation and Mission Design

AI is poised to influence not only forecasting, but also how space weather data are collected. AI-guided observing system simulation experiments can identify optimal sensor placements, constellation geometries, and measurement priorities to maximize forecast skill under budget constraints (McGranaghan et al., 2022). This feedback loop between prediction and observation design represents a fundamental shift from passive data usage to active, adaptive sensing. Future missions may incorporate onboard AI for event detection and adaptive sampling, reducing latency and improving coverage during rapidly evolving solar and heliospheric events.

10.5 Integration with Socio-Technical Decision Systems

Another critical direction is the deeper integration of space weather AI with socio-technical decision frameworks. Forecasts must be translated into actionable risk metrics tailored to specific sectors, such as power grids, satellite operations, aviation, and emergency management (Oughton et al., 2017). This requires interdisciplinary research at the interface of heliophysics, engineering, economics, and social science. In my assessment, the long-term value of AI in space weather will depend as much on how forecasts are used as on how accurately they are generated.

10.6 Governance, Open Science, and International Coordination

As AI-driven space weather systems grow in influence, issues of governance and equity will become increasingly salient. Future research must prioritize open data, transparent models, and reproducible benchmarks, enabling broad participation and independent verification (Schrijver et al., 2015; Eastwood et al., 2017). Given the global nature of space weather impacts, international coordination will be essential to ensure that AI-enabled advances contribute to shared resilience rather than fragmented or unequal outcomes.

10.7 Synthesis and Outlook

Future research in AI-based space weather prediction will be inherently hybrid, adaptive, and interdisciplinary. Advances in foundation models, rare-event learning, adaptive sensing, and digital twin architectures collectively point toward forecasting systems that are not only more accurate, but more scientifically insightful and societally relevant. The central challenge ahead is to ensure that these technological advances are guided by physical understanding, ethical governance, and a clear focus on public benefit.

12 Conclusions

Artificial intelligence has reached a level of maturity where it is no longer an experimental adjunct to space weather research, but a structural component of next-generation forecasting systems. Across solar, heliospheric, and geospace domains, AI-based methods have consistently demonstrated measurable improvements in forecast accuracy, lead time, and computational efficiency relative to traditional empirical and physics-only approaches. These gains are particularly evident in solar flare classification, CME arrival-time prediction, geomagnetic storm forecasting, and ionospheric disturbance modeling, where benchmark studies show reductions in error ranging from 20–60% depending on task and lead time (Bobra & Couvidat, 2015; Chen et al., 2019; Wintoft et al., 2017; Sun et al., 2020). At the same time, this review highlights that AI's true value lies not in replacing physical models, but in complementing and augmenting them. Hybrid AI–physics frameworks, physics-informed neural networks, and AI-initialized MHD simulations represent the most scientifically robust and operationally credible pathway forward. These approaches preserve causal structure and physical consistency while exploiting AI's capacity for pattern recognition, inference under uncertainty, and real-time emulation of complex systems (Camporeale, 2019; Karniadakis et al., 2021; Liemohn et al., 2020). The emergence of space weather digital twins marks a conceptual shift in the field—from episodic forecasting toward continuously evolving, closed-loop representations of the Sun–Earth system. By integrating real-time data assimilation, AI-based inference, and physics-based simulation within a unified architecture, digital twins enable anticipatory, scenario-driven forecasting and risk assessment. Such systems are poised to transform space weather services from reactive warning mechanisms into proactive tools for infrastructure resilience, mission planning, and societal risk governance (McGranaghan et al., 2022; Baker et al., 2018).

Nevertheless, significant challenges remain. Data sparsity for extreme events, non-stationarity across solar cycles, limited observability, model robustness under out-of-distribution conditions, and the need for explainability and trust in operational settings all constrain current capabilities. Without standardized benchmarks, transparent validation protocols, and sustained observational investment, AI-driven advances risk overfitting to historical conditions or eroding confidence among operational stakeholders (Schrijver et al., 2015; Pulkkinen et al., 2017). From a broader perspective, the integration of AI into space weather forecasting carries profound societal and economic implications. Improved forecasts directly support the protection of satellites, power grids, aviation, and communication systems, with potential to avert losses on the order of billions of dollars during severe events. Equally important are issues of governance, equity, and international coordination, as space weather impacts transcend national boundaries and disproportionately affect technologically dependent societies (Eastwood et al., 2017; Oughton et al., 2017). In closing, AI should be viewed not as a panacea, but as an epistemic accelerator—a tool that amplifies the reach of physical theory, enhances operational decision-making, and exposes gaps in our understanding of the coupled Sun–Earth system. The future of space weather prediction will be hybrid, explainable, uncertainty-aware, and human-centered. Realizing this vision will require sustained interdisciplinary collaboration across heliophysics, data science, engineering, and the social sciences, guided by a commitment to scientific rigor and public benefit.

Acknowledgement:

The authors gratefully acknowledge the World Data Centre for Geomagnetism, Kyoto, for providing definitive Dst and AE index data; the INTERMAGNET programme for INTERMAGNET data access; the OMNIWeb service of NASA GSFC for OMNI solar wind data; and the NOAA SWPC for GOES X-ray flux data. The authors declare no conflicts of interest. We are very much thankful to the authors of different publications as many new ideas are abstracted from them. Authors also express gratefulness to their colleagues and family members for their continuous help, inspirations, encouragement, and sacrifices without which this work could not be executed. Finally, the main target of this work will not be achieved unless it is used by research institutions, students, research scholars, and authors in their future works. The authors will remain ever grateful to Dr. Neelu Singh, Director, ICFRE Tropical Forest Research Institute, Jabalpur, Principal, Jabalpur Engineering College, Jabalpur & Principal Government Science College, Jabalpur who helped by giving constructive suggestions for this work. The authors are also responsible for any possible errors and shortcomings, if any in the paper, despite the best attempt to make it immaculate.

References:

[1] Agarwal R., Mishra Rajesh Kumar, (2013), 33rd International Cosmic Ray Conference (held in Rio de Janeiro in July 2013) At: Rio de Janeiro, pp. 25.

[2] Baker, D. N., Pulkkinen, A., Li, X., Kanekal, S. G., & McPherron, R. L. (2018). Severe space weather events—Understanding societal and economic impacts. Space Science Reviews, 214, 17.

[3] Bala, R., & Reiff, P. (2012). Improvements in short-term forecasting of geomagnetic activity. Space Weather, 10, S06003.

[4] Bobra, M. G., & Couvidat, S. (2015). Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. The Astrophysical Journal, 798, 135.

[5] Bussy-Virat, C. D., Oran, R., & Schrijver, C. J. (2018). Understanding the physics of extreme geomagnetic storms using machine learning. Space Weather, 16, 1891–1906.

[6] Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. Space Weather, 17, 1166–1207.

[7] Camporeale, E., Wing, S., & Johnson, J. R. (2021). Machine learning techniques for space weather forecasting. Frontiers in Astronomy and Space Sciences, 8, 639.

[8] Chen, Y., Manchester, W. B., Hero, A. O., Toth, G., & Lugaz, N. (2019). Predicting the arrival time of coronal mass ejections with machine learning. Space Weather, 17, 1404–1421.

[9] Conlon, P. A., Gallagher, P. T., McAteer, R. T. J., & Ireland, J. (2018). Automated detection of solar flares using SDO data. Solar Physics, 293, 55.

[10] Crown, M. D. (2012). Validation of the NOAA Space Weather Prediction Center's geomagnetic storm forecasts. Space Weather, 10, S06006.

[11] Deng, Y., Fuller-Rowell, T., & Chen, Y. (2021). Ionospheric total electron content prediction using deep learning. Journal of Geophysical Research: Space Physics, 126, e2020JA028689.

[12] Eastwood, J. P., Biffis, E., Hapgood, M. A., Green, L., Bisi, M. M., Bentley, R. D., … Gibbs, M. (2017). The economic impact of space weather: Where do we stand? Risk Analysis, 37, 206–218.

[13] Fox, N. J., Velli, M. C., Bale, S. D., Decker, R., Driesman, A., Howard, R. A., … Szabo, A. (2016). The Solar Probe Plus mission: Humanity's first visit to our star. Space Science Reviews, 204, 7–48.

[14] Glaessgen, E. H., & Stargel, D. S. (2012). The digital twin paradigm for future NASA and U.S. Air Force vehicles. Proceedings of the AIAA Conference, AIAA 2012-1818.

[15] Jaynes, A. N., Wing, S., & Kistler, L. M. (2020). The dynamic response of Earth's radiation belts to solar wind driving. Nature Physics, 16, 1046–1050. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3, 422–440.

[16] Kim, J., Kim, J. H., Moon, Y. J., & Park, E. (2021). Deep-learning-based geomagnetic storm prediction using solar wind parameters. IEEE Transactions on Geoscience and Remote Sensing, 59, 5233–5245.

[17] Leka, K. D., Barnes, G., Wagner, E. L., Yardley, S. L., Gary, G. A., & Ugarte-Urra, I. (2019). A comparison of flare forecasting methods using vector magnetic field data. The Astrophysical Journal Supplement Series, 243, 36.

[18] Liemohn, M. W., Ganushkina, N. Y., De Zeeuw, D. L., Welling, D. T., Jacobsen, K. S., & Pulkkinen, A. (2020). Uncertainty quantification and validation in space weather modeling. Space Weather, 18, e2019SW002372.

[19] Liu, H., Chen, Y., & Hu, Z. (2018). Machine learning-based prediction of geomagnetic storms. Space Weather, 16, 215–230.

[20] Lugaz, N., Manchester, W. B., & Gombosi, T. I. (2018). Numerical simulation

of coronal mass ejection propagation and interaction. Space Weather, 16, 37–62.

[21] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 4765–4774.

[22] Manchester, W. B., Kilpua, E. K. J., Liu, Y. D., Lugaz, N., Riley, P., Török, T., & Vršnak, B. (2017). The physical processes of CME/ICME evolution. Space Science Reviews, 212, 1159–1219.

[23] McGranaghan, R., Mannucci, A. J., Verkhoglyadova, O., Mays, M. L., Meng, X., & Burrell, A. G. (2017). Machine learning for solar wind and geomagnetic forecasting. Space Weather, 15, 122–134.

[24] McGranaghan, R., Zesta, E., & Connor, H. K. (2022). Toward a space weather digital twin. Space Weather, 20, e2021SW002856.

[25] Mishra, Rajesh Kumar, Mishra, Divyansh and Agarwal, R. (2024). Cosmic Rays and Space Weather Predictions: Exploring the Dynamic Universe, International Journal of Scientific Research & Engineering Trends, 10, 4, 1642-1672.

[26] Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., Ishii, M., & Shibata, K. (2018). Solar flare prediction model with three machine-learning algorithms using ultraviolet and vector magnetogram data. The Astrophysical Journal, 858, 113.

[27] Oughton, E. J., Skelton, A., Horne, R. B., Thomson, A. W. P., & Gaunt, C. T. (2017). Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. Risk Analysis, 37, 984–998.

[28] Owens, M. J., & Forsyth, R. J. (2013). The heliospheric magnetic field. Living Reviews in Solar Physics, 10, 5.

[29] Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. (2012). The Solar Dynamics Observatory (SDO). Solar Physics, 275, 3–15.

[30] Pulkkinen, A., Bernabeu, E., Eichner, J., Viljanen, A., & Ngwira, C. (2017). Regional-scale high-latitude extreme geomagnetic disturbance scenarios and their implications for electric power grids. Space Weather, 15, 828–856.

[31] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.

[32] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566, 195–204.

[33] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 1135–1144.

[34] Riley, P., Baker, D., Liu, Y. D., Verronen, P., Singer, H., & Güdel, M. (2018). Extreme space weather events: From cradle to grave. Space Weather, 16, 1245–1267.

[35] Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., … Bengio, Y. (2019). Tackling climate change with machine learning. arXiv preprint, arXiv:1906.05433.

[36] Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (2021). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer Nature, Cham.

[37] Schrijver, C. J., Kauristie, K., Aylward, A. D., Denardini, C. M., Gibson, S. E., Glover, A., … Walsh, B. (2015). Understanding space weather to shield society: A global road map for 2015–2025. Advances in Space Research, 55, 2745–2807.

[38] Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., & Snow, F. (1998). The Advanced Composition Explorer. Space Science Reviews, 86, 1–22.

[39] Sudar, D., Vršnak, B., Dumbović, M., & Temmer, M. (2020). Machine-learning approach to solar flare prediction. Solar Physics, 295, 37.

[40] Sun, W., Reichstein, M., & Brenning, A. (2020). Ionospheric total electron content forecasting using deep learning. Journal of Geophysical Research: Space Physics, 125, e2019JA027684.

[41] Temerin, M., & Li, X. (2006). Dst model for 1995–2002. Journal of

Geophysical Research: Space Physics, 111, A04221.

[42] Tóth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., de Zeeuw, D. L., … Kóta, J. (2005). Space Weather Modeling Framework: A new tool for the space science community. Journal of Geophysical Research: Space Physics, 110, A12226.

[43] Tuegel, E. J., Ingraffea, A. R., Eason, T. G., & Spottswood, S. M. (2011). Reengineering aircraft structural life prediction using a digital twin. Proceedings of the AIAA Conference, AIAA 2011-1818.

[44] Vourlidas, A., Patsourakos, S., & Savani, N. P. (2019). Predicting the geoeffectiveness of coronal mass ejections: Current status, open issues, and path forward. Space Science Reviews, 215, 38.

[45] Wiener, J., Lugaz, N., & Manchester, W. B. (2020). Machine learning for predicting the arrival time of interplanetary coronal mass ejections. Space Weather, 18, e2019SW002455.

[46] Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating physics-based modeling with machine learning: A survey. Proceedings of the National Academy of Sciences, 119, e2109730119.

[47] Wintoft, P., Wik, M., Matzka, J., & Shprits, Y. (2017). Forecasting Kp index using LSTM neural networks. Journal of Space Weather and Space Climate, 7, A24.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.