BILINGUAL PUBLISHING GROUP

Journal of Environmental & Earth Sciences
https://journals.bilpubgroup.com/index.php/jees

REVIEW

From Satellites to Sensors: Harnessing AI to Unify Multi-Scale Data in Modern Atmospheric Monitoring

Yan Wu

School of Electronics and Information Engineering, Liaoning Institute of Technology, Benxi 117004, China

ABSTRACT

Software-defined, data-intensive cyber-physical systems and software-defined networks of atmospheric observers are evolving rapidly due to the rapid expansion of sensing diversity, the volume of streaming data, and the demand for low-latency, decision-relevant products. Simultaneously, artificial intelligence (AI) and the continuously evolving state of computing are making it possible to create end-to-end architecture fostering the migrations of the presumably single algorithm to combined intelligent ingestion, quality control, and multi-modal fusion, uncertainty-related retrieval, and scalable service delivery at the edge-to-cloud-high-performance computing (HPC) environment. This overview summarizes AI-based models of future atmospheric observation networks within a single, consolidated taxonomy based on deployment topology, learning and update modes, connectivity to physical models and data assimilation, level of autonomy (passive to adaptive sensing), and model of governance. Next, we consider recurring architectural themes, such as edge intelligence and streaming provenance and machine learning operations (MLOps)/model operations (ModelOps) to continue evaluation and safely update, and we scrutinize integration gateways with physical models, like data-assimilation-oriented outputs, hybrid/physics-informed designs, and simulation of observing systems using digital twins. Lastly, we address evaluation and readiness aspects that are not limited to predictive skill, but also involve calibrated uncertainty, nonstationary and extreme robustness, system latency and reliability, interoperability, security, and demonstrated downstream influence on analyses and forecasts. Through bringing together the cross-cutting issues and prospects, this review provides a road map with respect to trustworthy, interoperable, and sustainable observation infrastructures in which code and climate science

*CORRESPONDING AUTHOR:
Yan Wu, School of Electronics and Information Engineering, Liaoning Institute of Technology, Benxi 117004, China; Email: wuyan@lnist.edu.cn

ARTICLE INFO
Received: 12 December 2025 | Revised: 1 February 2026 | Accepted: 4 February 2026 | Published Online: 5 February 2026
DOI: https://doi.org/10.30564/jees.v8i2.12929

CITATION
Wu, Y., 2026. From Satellites to Sensors: Harnessing AI to Unify Multi-Scale Data in Modern Atmospheric Monitoring. Journal of Environmental & Earth Sciences. 8(2): 72–104. DOI: https://doi.org/10.30564/jees.v8i2.12929

will co-evolve.

Keywords: Atmospheric Observation Networks; Data Assimilation; Edge AI; Uncertainty Quantification; Digital Twins

1. Introduction

The network of atmospheric observation is at the frontline of climatic and weather intelligence. They furnish measures on which day-to-day forecasting, hazard-early warning, air-quality guidance, and climate forecasting far in the future are extensively dependent. However, the business environment within which these networks operate has evolved radically. The pressures on observation systems caused by urbanization and increasing exposure to extreme events, and increasing demands on multi-resolution decision support, are compelling observation systems not only to provide more data but to provide more actionable data at a higher rate, with a quantified level of uncertainty, and unique formats compatible with immediate use in downstream models and services [1,2]. Simultaneously, there is diversification of sensing technology: alongside the conventional in situ stations and radiosondes are now expansive radar and satellite constellations, aircraft measurements, Internet-of-Things (IoT) sensors of small size, customer-science feeds, and mobile applications. The resulting topography is uneven in terms of sampling frequency, representativeness of space, stability of calibration, and error properties. Concisely, the observation network has emerged as an intricate cyber-physical network, the functionality of which relies equally upon software and data pipelines as much as on the sensor hardware directly [3,4].

This change overlaps a second change; the recent progress in artificial intelligence (AI) and current software engineering has enabled extracting value out of atmospheric observations in previously impractical ways [5]. Deep learning has made it possible to combine multi-modal data, do gap-filling, anomaly detection, and latent variable inference using remotely sensed signals. Self-supervised and foundation-model methods have improved the re-use of representations across tasks and scale, making it possible to have models trained on a massive archive and extrapolate to new areas and sensors. In the meantime, cloud-native infrastructure, stream computing, and edge computing have facilitated the continuous inference near sensors, shrinking the latency and bandwidth requirements. Collectively, these trends are transforming the main question: How do we get more observations into more reliable, interoperable, and decision-ready products? How do we create AI-based frameworks that can consistently convert raw observations into trustworthy, interoperable, and decision-ready products?

More importantly, the solution is not simply using AI. In atmospheric science, the usefulness of any measure is achieved by a sequence of procedures, ingestion, quality control (QC), calibration, spatiotemporal fusion, inter-instrumental fusion, uncertainty quantification, physical modeling assimilation, and providing it back into forecasting or decision processes [6]. Weak links at any point of this chain will vitiate the results downstream. To give an example, a small bias of a cheap sensor network can cause systematic errors in the fields being derived; a slow delivery of otherwise accurate measurements can diminish the impact of a forecast; and falsely-specified uncertainties may cause over-confident products at the point of greatest need, at the extremes. It is possible to assist in every step using AI, though, only when it is put into resilient architectures, which handle data provenance, model versioning, drift control, security, and operational limits. Therefore, this review centers on the intersection of code and climate: the advent of the deployable end-to-end systems that relate the atmospheric sensing to both the scientific and societal worth [7,8].

The word framework is not used by chance. Most of the contributions to the literature focus on a particular model (e.g., a neural network for radar nowcasting) or perceiving a single functionality (e.g., anomaly detection for QC). These advances are valuable, yet they do not typically give attention to the system-level requirements that define whether or not an approach can be believed in and supported in the operational environment. In comparison, an AI-driven framework has a more extensive scope: (i) ingestion and data management software; (ii) one (or more) learning or inference software (QC) systems; (iii) deployment patterns (i.e., edge devices, cloud services, and high-performance computers); (iv) mechanisms to secure, privacy, and control access; and (v) lifecycle practices (such as MLOps/ModelOps) to constantly evaluate and update. This broader definition is consistent with how next-generation observation systems are literally constructed and operated, and it can be used to understand why two procedures with comparable predictive ability may have such a very different impact in reality [5,9,10].

Another important driving force of the framework perspective is the emergence of real-time and high-density observing. Conventional networks were modeled based on the periodic transmission and centralized processing, where the assimilation cycles and product generation occur at a fixed point. Conversely, new deployments are more and more implemented in the form of streaming systems. Dense sensor arrays produce continuous flows that need to be filtered, summarized, and prioritized. High-volume gridded products are offered by satellite and radars, and irregular trajectories are offered by mobile platforms. With this kind of regime, one centralized pipeline will turn into a bottleneck. Running inference close to the sensor, then known as edge computing, has consequently been appealing to on-device QC, event detection (e.g., rapidly growing signatures of convection), adaptive sampling triggering, and intelligent compression. Edge deployment, however, brings about additional limitations such as insufficient power, intermittency, heterogeneity of hardware, thermal constraints, and robust update mechanisms are required. It also poses new scientific challenges: How will it be possible to trace uncertainties when the section of the pipeline functions not in a centralized way? How should the division of labor between edge and cloud be done? What to do to make models stable in both distribution shift and sensor drift? These are the questions that cannot be answered only through the selection of algorithms; these are the questions that require architectural and governance answers [11–13].

The other force is the growing enthusiasm for the kind of closed-loop observation system, in which sensing is not necessarily passive. In closed-loop systems, observations at a point and in time are informed by forecasts and uncertainty by the use of a model that finds extensive application in the field of targeted observations, but is now many times more within the reach of automation and AI. Limited sensing resources can be allocated, transmissions can be prioritized, or the mobile sensors can be directed to high information gain regions using active learning and reinforcement learning. The simulators that model both sensor behaviors and atmospheric processes, when combined with an observational network, are known as the digital twins of these networks, and they provide a way to experiment with such strategies by simulating experiments of observational systems (OSSEs). However, closed-loop systems increase the requirements on trust: defective targeting policies may systematically ignore crucial conditions; adversarial or corrupt inputs may confuse sensing, and defective uncertainty calibration errors may lead the system to seek spurious patterns. Fail-safe modes, monitoring, human-in-the-loop controls, and auditability are provided at the framework level and hence are necessary [14,15].

Physical modeling combines with AI, which adds more complexity. The observation networks in the atmosphere are closely integrated with physics-based numerical weather prediction (NWP) and reanalysis systems in the form of data assimilation (DA). DA is a principled approach to integrating observations with model forecasts of the usage of correctly specified observation operators and error models. Here, AI could serve in several ways, especially approximating costly operators that can be learned fast, biases can be reduced, errors in observation approximated, or the subgrid processes that influence representativeness can be modeled. Combining physical constraints with data-driven learning, Hybrid methods, a combination of both, are potentially useful towards enhancing generalization and consistency. But these solutions generate new integration problems: how to make learned components inter-regime stable, how to measure the uncertainty in learned operators, and how to design interfaces such that the outputs of AI systems can be safely consumed by a DA system. Consequently, the range of coupling between AI structures and physical models, including loosely coupled analytics pipelines, DA-based structures, and fully digital-twin ones, is in the list of the most significant themes of this review [16].

Although at a high pace of development, there are a number of enduring challenges that restrain the translation of research prototypes to sustainable infrastructures. To begin with, interoperability is not even. The information stored in observation is often spread out in formats and metadata standards, and AI-ready datasets also record a large amount of information in formats that are not reproducible or uniform. Second, the practices of evaluation are piecemeal. Lots of studies report model skill measures and fail to measure the impact of the forecast, its latency, ability to remain robust in the presence of a shift, or operational maintenance load.

Third, the problem of governance and security is not well covered in much of the scientific literature. Networks based on observation are becoming susceptible to spoofing, tampering, and supply-chain risks, particularly where they use low-cost or distributed sensors. Fourth, equality and coverage are important issues: AI-driven optimization may entail the absence of incentives toward areas with high infrastructure and past data, which serves as an improvement of the observational gaps in underserved regions. Lastly, the issue of sustainability has been made a first-order agenda item. Training large models should be weighed against the costs of training and energy, and the cost of continuous inference needs to be mitigated against the results of better products, and the lifecycle carbon effects should have been considered when designing systems[17,18].

The reason behind this review is the necessity to accord some sanity to this rapidly changing space. We construct AI-based observational networks of the atmosphere based on a single taxonomy grounded on a comparative table that characterizes representative systems based on their layer of deployment (edge/cloud/hybrid), learning paradigm (supervised, self-supervised, foundation, online), strength of coupling between them and their physical models and DA (standalone to integrated), control (passive/adaptive observing), and governance model (centralized to federated/privacy preserving). Developing this taxonomy, we examine end-to-end architectural patterns, general ability modules (QC, fusion, retrievals, nowcasting, uncertainty quantification—UQ), and commentary on the integration of these general competencies and approach to integration with DA and digital-twin ecosystems. We further suggest an evaluation and preparation optic, which extends the predictive capability to encompass latency, dependability, reproducibility, security stance, sustainability, and effects on downstream projections and forecasts. Lastly, we bring together cross-cutting constraints and future directions and highlight interoperability principles, enduring MLOps + DA pipelines, reliable uncertainty about extremes, collaborative federation, benchmark suites to connect sensing decisions with effect prediction, and the energy-aware design[19–21].

The remaining part of the paper has the following structure. Section 2 presents a background and abstract principles of how the structure of errors, concept of data fusion, and theory of data assimilation are presented that form the stratosphere of AI integration. Section 3 gives the taxonomy and the table of overview of AI-driven structures, which gives a recorded landscape map. In Section 4, there is a discussion on end-to-end its architecture and core capabilities in edge, cloud, and HPC deployments. Section 5 aims at physical modeling, coupling, data assimilation, and adaptive observing control, such as digital twins. Section 6 deals with evaluation, operation readiness, cross-cut challenges, and future directions. Section 7 has its conclusions and a few insights into deepening what it will require to develop next-generation, closed-loop, reliable infrastructures of atmospheric observation at the tightest coordinated code and climate science co-evolution[22].

2. Background and Theoretical Foundations

Next-generation networks Observations of the atmosphere Generation Next-generation memory systems are at the interface of the geophysical measurement science, statistical inference, and large software systems. It must also be evidently theoretically grounded in understanding what an AI-driven framework adds to the traditional pipelines and in determining where AI can be implemented without breaking physical consistency, the requirement of uncertainty, or the operational constraints. Highlighted in this section are the observation modalities and error structures defining atmospheric data sets, the essential concepts of data quality control and multi-sensor fusion, and the concepts of uncertainty quantification and data assimilation form the basis of contemporary forecasting and reanalysis. These concepts give the reference points to be employed in the remainder of the taxonomy and comparison in the following parts[23,24].

2.1. Atmospheric Observation Networks as Cyber–Physical Systems

Atmospheric observing systems should be considered as cyberphysical networks: sensors are constructed physically, and through software pipelines, the signals are processed into geophysical variables and products of the solution. The stations fixed on site include applications of meteorological (and air-quality) and upper-air (radiosondes and profilers), remote sensing (radar, satellite radiometers and sounders, lidar), aircraft and ship measurements, and a growing ecosystem of low-cost IoT and mobile sensors. The samples of the variables, scales, and regimes of each modality differ, with each having different latency and coverage patterns[25].

Systemically, the data flow within the system functionality of the observation pipeline may have (i) sensing and on-board preprocessing, (ii) transmission and ingestion, (iii) calibration and quality control, (iv) spatiotemporal harmonization, (v) high-level retrieval or feature generation, (vi) fusion and uncertainty characterization and (vii) downstream connection to models, assimilation and applications. The operation of next-generation networks focuses more on streaming, heterogeneous compute (edge/cloud/HPC), and constant refresh of both data and models, which further drives the requirement of explicit theory-based design, as opposed to algorithm insertions[26].

2.2. Measurement Error, Bias, and Representativeness

The usefulness of an observation is not only offset by its nominal accuracy, but also by its error model of the behavior of uncertainty in time, space, and operating conditions. Error in the dataset in the atmosphere is hardly independent and identically distributed. Rather, they may contain random noise, systematic bias, drift, change of continuity at a point of maintenance or sensor replacement, and correlated errors caused by exposure to environmental impacts or common processing processes. One is representativeness error (the discrepancy between what a sensor captures (ex, at a point or along a path), and what a grid-based model state variable specifies (an area volume mean at finite resolution). The importance of representativeness error is majorly on large urban networks in dense urban settings, highly varied terrain, convective storms, and near-source air-quality measurements, where the gradient and variability across grids are very slight. In the case of AI-based frameworks, representativeness error occurs in training labeling, the definition of ground truth, and the consistency of learned mappings to scale variations. It provides motivation as well as to hybrid methods that directly model scale conversion, use of physical constraints, or learn context-specific observation uncertainties[2,27].

Prejudice is also underlying. Most types of sensors are subject to state-dependent bias (e.g., humidity sensor errors with temperature), regime-dependent behavior (e.g., precipitation effects on measurement), and processing bias (e.g., artifacts of retrieval algorithms). Since data assimilation and probabilistic prediction are extremely sensitive to bias and mis-specified uncertainty, such a claim purporting operational benefit should focus on such corrections as well, bias correction and uncertainty calibration, preferably with transparent provenance and version control[28].

2.3. Data Quality Control Foundations

The entry point of observation utility is quality control (QC). In classical QC, instrument diagnostics, range checks, temporal/spatial consistency checks, and buddy checks (requests to other sensors in the field or other background fields) are used together. Though all these means are still instrumental, the subcutaneous Fleet to heterogeneous, high-density networks introduces new demands on QC: recognition of subtle drift, detecting sensor malfunction in infrequent cases and intermittent connectivity, and missingness. AI techniques, e.g., anomaly detection, sequence models to ensure temporal consistency, and graph-based models to ensure neighborhood consistency, are now more frequently used to supplement rule-based QC. The theoretical basis is the same: QC is an uncertain problem of decision. It should trade off between false positives (rejecting useful data) and false negatives (accepting injurious data), and it should maintain a record of provenance such that users of that data can see why data were de-junked, fixed, or marked down[29].

A strong structure, thus, is advantageous for clear QC outputs other than a pass/fail indication. Some intermediate products that can be useful are QC flags where semantic categories are, estimated bias/drift elements, uncertainty inflation elements, and underestimation of confidence, which can be reused in fusion and assimilation. QC is also related to governance: in case the network is federated or has community sensors, decisions about QC can be decentralized, which demands uniform standards, audit logs, and security controls[30].

2.4. Multi-Sensor Fusion and Spatiotemporal Harmonization

It is observed that increasing numbers of observation networks are based upon fusion: a combination of measurements of a variety of platforms to generate fields of better coverage, resolution, or variable completeness. Fusion is diffused by different sampling geometries (point vs. volume), varying temporal cadence, noise, and bias, as well as inconsistent metadata. Approaches with theoretical underpinnings are interpolation and kriging, Bayesian hierarchical modeling, variational formulations, ensemble approaches, and contemporary learning-based spatiotemporal modeling. Before any fusion, coordinate systems, time stamps, vertical references, units, and metadata must be harmonized and known biases corrected, and data gaps taken into consideration. In the case of learning-based fusion, harmonization additionally involves building consistent training objectives, loss functions are chosen indicating uncertainty and scale, and leakage needs to be avoided (e.g., being trained on information that is inapplicable to a real-time setting)[31].

One concept of fusion that is important to atmospheric networks is information complementarity. Radar can give high-frequency precap structure but low microphysical specificity; satellites can give low-frequency coverage but have definable ambiguities in retailing; and in situ sensors can give low-frequency conditions at the near-surface but can provide little coverage between the surface and the atmosphere. Well-organized structures take advantage of complementary strengths by having architectures that are capable of accepting multi-modal inputs and indicating uncertainty in the event of modalities being inconsistent or unavailable.

2.5. Uncertainty Quantification and Calibration

Decision-making in the atmosphere is probabilistic in nature. Not only best predictions but plausible uncertainty, such as probabilistic hazard advice against heavy paper, wind, heat stress, or air-quality supersenses, can be needed by its users. Since AI components start to be put in the observation pipelines, quantification of uncertainty (UQ) turns into a first-order requirement, since when the learned models are used, they may be overconfident, particularly in the case of distribution shift or rare extremes. Theoretically, UQ consists of (i) aleatoric uncertainty (irreducible measurements and varied environment), and (ii) epistemic uncertainty (model uncertainty created out of forced scarcity of data and imperfect structure or out-of-distribution), available to us. Such practical UQ mechanisms as ensembles, Bayesian approximations, quantile regression, conformal prediction, and calibrated probabilistic outputs can be identified. Calibration is essential: a probabilistic product needs to be reliable, i.e., the predicted probabilities need to match observed frequencies, and has to be sharp enough to be of use[32,33].

UQ should also be computationally viable when it comes to edge and cloud frameworks. The edge inference might need lightweight approximations or compressed ensembles, whereas cloud/HPC techniques might be able to access heavier probabilistic terms. Whichever gets implemented, one has to monitor uncertainty throughout the pipeline, QC, fusion, retrieval, and assimilation, to have downstream systems weighing observations accordingly.

Aleatoric and epistemic uncertainty should be used differently to facilitate the operational use of AI-driven observation pipelines. Aleatoric uncertainty is due to variations in the atmospheric processes and noise of measurements themselves, whereas epistemic uncertainty is due to insufficiencies in model structure, the extent of training data, or representation learning[34,35]. These two types of uncertainty have different effects on the decisions made in an operational setting: aleatoric uncertainty limits the amount of predictive precision that is possible, and epistemic uncertainty can be minimized with the help of extra observations or by refining the model. The modern techniques of estimating uncertainty are ensemble learning, Bayesian approximations, quantile regression, and conformal prediction[36]. Nevertheless, its implementation in real-world edge settings is difficult due to resource constraints in those settings, where computational constraints limit the size of an ensemble to utilize Bayesian inference. In turn, lightweight uncertainty surrogates and predictive intervals that are calibrated are still an active line of research.

2.6. Data Assimilation Essentials for Framework Integration

The formal apparatus of assimilation between observations and dynamical model forecasts to estimate the atmospheric state is offered by data assimilation (DA). Although there are many variations of DA methods (e.g., variational, ensemble-based, hybrid), they have some similar ingredients: an estimate of the background (prior), an operator of observation of model state to the space of observation, and error statistics of both the background and the observations.

This is critical in terms of attitude to bias, representativeness, and well-specified uncertainties. In the case of AI-driven frameworks, DA poses opportunities and limitations. AI may enhance observation operator (e.g., learned retrievals or surrogates), provide approximations of observation error, and/or correct bias. Nevertheless, any learnt component should be in line with DA assumptions and stability demands. Indicatively, a time-dependent learned bias correction has to be versioned and regime-switching checked; a learned operator has to maintain necessary monotonicity or physical conditions where needed; and uncertainty estimates must be plausible because misrepresentation of observation errors can cause sharp deteriorations in the performance of a DA[37,38].

DA also encourages evaluation techniques that are different than independent ML metrics. A processing improvement in observation is useful when it enhances the accuracy of analysis or forecast skill, minimizes systematic errors, or enhances reliability, which is often assessed by observing system experiments (OSEs) and observing system simulation experiments (OSSEs). These evaluations of the DA form the base of further discussion of the readiness and impact in the review.

2.7. Software and Interoperability Foundations

Since observation networks are operational structures, software engineering principles are not secondary—they are also among the theoretical foundations of implementable systems. In essence, it is required to be reproducible (has versioning with data and models), name-traceable (provenance and audit logs), reliable (monitored), failover-resilient, and rollback-capable, and be interoperable (matamata-minimal patterns and application program portability). This is what these requirements look like: models need to be testable: their inputs are observable, updatable, and verifiable in the production process, and their outputs need to be interpretable by a downstream machine of production, like DA and decision services. Interoperability is of great concern, especially in the multi-agency and cross-border setting. Architects integrating AI into observation curves have to be consistent with preexisting communal customs concerning metadata, quality flags, and provenance, or the advantages of enhanced analytics, questioning of integration drag. Interoperability and lifecycle management are thus considered as baseline in this review, on the same level as learning algorithms and physical coupling[39].

The above theoretical grounds explain why AI-powered observation systems should be considered in the form of end-to-end form. Complex, regime-dependent bias and uncertainty in atmospheric observations; need to be brought into harmony through fusion and propagation of uncertainty in a principled manner; and utilize rigid demands on operators of observation and error statistics in DA. In turn, successful frameworks in next-generation networks are likely to have a number of common principles: they express uncertainty, maintain provenance, scale across representativeness, support streaming and heterogeneous compute, and offer interfaces that can be successfully fused and assimilated[40].

These concepts drive the taxonomy presented in Section 3, in which the frameworks are grouped by the deployment layer, learning paradigm, strength of coupling with physical models and assimilation, level of adaptive control, and the governance model. They also provide information to the lens of evaluation in Section 6, which focuses not just on predictive ability, but also on calibration, ability to survive a shift, latency, and reliability, as well as the shown effect on downstream analysis and forecasting.

3. Taxonomy of AI-Driven Frameworks for Atmospheric Observation Networks

It has developed fast in the use of artificial intelligence to observe the atmosphere, yet the literature on this area is hard to compare due to the frequent use of similar words to describe radically different system realities. Other studies are based on one model that enhances a particular task like radar-based nowcasting, satellite retrieval, or anomaly detection. Additional contributions explain deployable pipelines which combine ingestion, quality control, uncertainty management and delivery of products to active users or to data assimilation systems. Since it is observed in practice, next-generation observation networks are becoming more of a cyber-physical infrastructure than a standalone instrument, this review uses a framework-based perspective. A framework here is considered to be a conglomeration of socio-technical end-to-end systems that involves software architecture, AI approaches, data governance and deployment plan to transform the raw observations to investigable, interoperable outputs within defined operational boundaries[3].

This taxonomy aims to supply a standardized set of terms that can be used to state what these structures are, how they are developed and how they are to be tested. The taxonomy should be laid out in such a way that it directly shows unexpectedly heterogeneous contributions in the comparative table and that it facilitates synthesis between otherwise extremely heterogeneous contributions. It also aims to relate the selection of a given architecture and learning paradigm to the nature of uncertainties, failure modes, and evidence of proof of its validity that suit a given framework. It is important in practice that the taxonomy should have continuous spectra instead of discrete bins, but explicit axes, too, are useful since they reveal the trade-offs and the gaps that are underutilized in the extant research and deployments[41].

3.1. Defining "Framework" in the Context of Atmospheric Observation Networks

In atmospheric science, the value of an observation is realized through a chain of transformations that begin at the sensor and extend to downstream models and decisions. A framework therefore, encompasses more than a predictive model. It includes the mechanisms by which data are acquired and ingested, validated and corrected, fused across modalities, annotated with uncertainty and provenance, and delivered in forms that downstream systems can reliably consume. It also includes lifecycle practices that determine whether the system can remain trustworthy over time, such as version control for datasets and models, monitoring for sensor drift and distribution shift, security hardening, and procedures for updates and rollback[1].

This definition reflects two realities. First, the operational requirements of observation networks—latency targets, bandwidth limitations, intermittently connected devices, safety considerations, and regulatory or privacy constraints—strongly shape what AI can do and how it must behave. Second, the downstream value of observations frequently depends on compatibility with data assimilation and physical modeling workflows, which impose constraints on bias, representativeness, and uncertainty specification. An algorithmic taxonomy will therefore miss the system-level characteristics that determine impact. For this reason, the taxonomy proposed here is structured around axes that capture deployment, learning, and update mode, physical-model coupling, autonomy, and governance[42].

3.2. Rationale for a Multi-Axis Taxonomy

The importance of a multi-axis taxonomy is due to the increased heterogeneity in atmospheric observation networks in both sensing and computation. An embedded pipeline creating QC flags on a low power gateway and only emitting them can be labeled the same AI-enabled observation system, or alternatively a cloud-native platform merging satellite, radar, and in situ data into real real-time gridded product, or as an assimilation-integrated workflow proving useful increases in forecast skill. Unless these differences are systematically described, comparisons will be misleading, and review conclusions will run a risk of being able to put different issues of concern into one narrative[43].

Another reason is that expectations of validation vary according to the type of system. A framework that generates stand-alone products can possibly be checked by predictive accuracy, calibration, and perturbation robustness. Conversely, a model aimed at facilitating assimilation has to be evaluated with regard to analysis/forecast effect and stability in an integrated modeling system. Likewise, adaptive sensing frameworks or closed-loop controls trigger decision-theoretic risks, which necessitate extra safety and audit, and human safety measures. To an Science Citation Index (SCI)-standard synthesis, a taxonomy relating the purpose of the system and the amount of its coupling with its evaluation criteria is thus crucial[44].

3.3. Axis A: Deployment Layer and Execution Topology

The former axis defines the location of the core of the intelligence and the processing of the framework. And this is not just a technical aspect of engineering; it actually dictates what information can be available at the time of inference, what complexity of models is viable, what is the feasible bandwidth and latency, and what types of monitoring and governance can be approximately imposed. In edge-deployed architectures, inference is deployed near the sensors, embedded devices, field gateways, or station controllers. They are frequently designed to address the requirement of a quick QC, event notification, smart compression, or priority transmission in case of intermittent connectivity. On the edge systems, power, compute, and memory restrictions are very strict. They also have practical needs like the ability to have reliable model updates and fail-safe behavior, and corrupted or spoofed inputs in uncontrolled environments[45].

Cloud or data-center architectures are centrally executed and are usually optimized in terms of scalability, availability of integration between data providers, and ongoing product delivery through APIs. They are inherently appropriate to multi-modal fusion, gap-filling, and near-real-time gridded products, which need the ability to access the large spatial context and massive training data. They additionally support fully-grown MLOps cultures, including evaluation pipelines and central monitoring boards, but need to deal with costs, energy usage, and dependency complexity.

Through HPC and operational modeling spaces, they have frameworks that are closely integrated with numeric-based models and assimilation cycles. In this case, the framework can use learned operators of observation, bias correction, uncertainty modeling, or a combination of the terms of hybrid correction, which should be stable to repeated cycling and consistent with the assumption of assimilation. HPC environment can sustain computationally intensive approaches, and uncertainty approaches based on ensembles; however, it is among the systems with high reproducibility and configuration checks since minimal modifications can spread across found compositions[46].

Chessboard role sedge, cloud, and HPC combine deliberately. As an example, edge components can do first pass QC and summarization, cloud components can do fusion and serve products, and HPC components can assess forecast impact or perform assimilation experiments. Hybrid patterns are becoming more popular as they provide a principled means of trading off between latency and bandwidth and context-rich inference and impact validation. They are, however, also more complex to integrate and need a careful design of interfaces in such a way that uncertainty, provenance, and versioning are consistent across layers[47].

3.4. Axis B: Learning Paradigm and Update Mode

The second axis outlines the process of training, adapting, and maintaining the AI components of the framework in the long term. This is closely interconnected with sensor drift, regime shift, and changing network structure in atmospheric observation networks. Paradigm of learning has an impact on data requirements, generalization behavior, and operational risk. Under supervised learning, whose targets are provided by curation (e.g., labelled QC flags), a retrieval training set based on physical inversion procedures, and pseudo-labels based on reanalysis, is still common. Its advantages are that it is easy to evaluate and interpret the error measures, but it is brittle when the labels are biased, not complete, or not representative of the condition of operation, particularly at extremes[48].

Representation learning and self-supervised learning use any available huge collections of unlabeled observations to learn transferable features. This paradigm is desirable in atmospheric observation, since high-volume data are available and dependable labels are frequently not available and costly. Such representations that are self-supervised will be able to enhance location and sensor diversity, yet they will still rely on downstream tasks to be calibrated, and the permanence of uncertainty quantified across shifts.

These thought trends are extended to foundation-model models, which are trained on a large scale, possibly in a variety of modalities and tasks, and tuned to particular observation networks during fine-tuning or lightweight adapters. Portability and the necessity to use task-specific labeling are the advantages of foundation models that are especially applicable in rapidly changing sensor ecosystems. The difficulty is that general representations have to be accompanied by stringent measurements, calibration, and domain adaptation proposals since a very powerful model still turns into a reliable falsehood when facing an undiscovered regime or unseen instrument artifacts[49].

Online and continuous learning characterize models that are updated at run time, whether by streaming information, by retraining periodically, or by automatically recalibrating. These measures are convincing in terms of correcting the drift and developing networks, which impose a major level of governance and safety concerns. Any update mechanism within the operational observation networks must be auditable, revertible, and observed to have no unintended performance regressions. Approaching ontology, the update mode is picked up as a first-class attribute of taxonomy: online learning frameworks are qualitatively different than those that are frozen at deployment, and support assessment should involve stability and rollback capability.

3.5. Axis C: Degree of Coupling to Physical Models and Data Assimilation

The axis is the directness of the interaction between the framework and physics-based models and assimilation workflows. The issue of coupling is that assimilation is not merely another consumer of data, but comes with very strict requirements concerning the bias management, representativeness, and the specification of uncertainty. In addition, the final social utility of an observation network can be in the form of better analyses and better forecasts, rather than sole measures. Isolated analytics systems can generate QC decisions, fused products, or retrievals, or nowcasts, which can be useful in their own right, especially in terms of monitoring and quick action. Predictive skill, calibration, robustness, and latency may be used to evaluate these systems, although they may not have a specific demonstration of the impact of the forecast[50].

The output of DA-aware frameworks is specifically structured such that they can be assimilated into an environment even when they do not form part of the DA system. Such would be structures that give out observation-error estimates, bias-corrected variables, or learned observation operators that are to substitute or supplement classical retrieval chains. The design of DAs needs to pay tight attention to the calibration of uncertainty and provenance recording since even minor mis-specifications are exaggerated by assimilation systems. DA-integrated paradigms incorporate elements in the assimilation cycle and assess them in terms of their impact on the analyses and forecasts. This type consists of systems that exhibit impact through system experiment observation or cycling studies of assimilation. DA-integrated methods are the best supported methods that are scientifically valuable and require the most engineering rigor since they need to behave consistently over repeated cycles and survive changes in operational regimes[51].

Digital-twin and simulation-coupled systems go further and couple a model of the observation network itself, including, frequently, an atmospheric simulator. Among the inquiries that digital twins will support in the design and policy include the location of the sensors, maintenance, and adaptive sampling. Their plausibility is determined by the faithfulness of reproducing sensor behavior, environmental limitations, and pipeline delays into the simulator. Since they can be employed to make closed-loop decisions, they need to be given the same serious attention as coupled modeling systems, with due consideration to uncertainty and reality testing.

3.6. Axis D: Autonomy and Control in the Observation Process

The fourth axis differentiates between those frameworks that process the observations on their own or have a direction on the sensing behavior. This is more so because the observational networks are becoming dynamic and resource-conscious. The passive structures handle any information that the network can give and concentrate on QC, fusion, retrieval, and delivery of products. Their risk is mainly connected with data resilience, uncertainty, and resilience in shift, and their operational protection is usually presented by fallback modes and monitoring. Assistive-control structures make no direct transformations of sensing but make recommendations to enable transmission prioritization, maintenance notifications, or recommended sampling volume adjustments. These systems introduce considerations of decision support and involve revealing information on confidence and failure modes, but human operators usually have the last word[52].

Adaptive-sensing architectures apply closed-loop control by stimulating measurements, maneuvering mobile platforms, distributing bandwidth, or making observations towards those areas with maximum information value. This autonomy can bring efficiency in information and enhance performance when the event is high-impact, but also new failure modes are generated. The objective function might have biases that result in systematic under-service because a region can be underserved, uncertainty might be inefficiently set, resulting in overconfident targeting, and adversarial manipulation can shift sensing resources. Due to these facts, the degree of autonomy should be clearly given in the taxonomy, and adaptive structures should comply with more rigorous requirements of safety restrictions, auditability, and human override[53].

3.7. Axis E: Governance, Privacy, and Data-Sharing Model

The fifth axis explains the manner in which data and models are governed in terms of organizations, jurisdictions, and stakeholders. Governance is not there in an ancillary manner: it defines what data can be used to actually develop training, what it is possible to evaluate, and what mechanisms exist to carry out accountability in a situation where something has gone wrong. Centralized governance presupposes the aggregation of data of a single operator or a closely coordinated consortium. It allows standardization, centralization of monitoring, and uniform update policies; however, in infeasible in circumstances when ownership of data is dispersed, or privacy and sovereignty hold weight.

Federated and distributed governance is reminiscent of multi-owner environments where raw data cannot be centralized. Frameworks in this category can either be federated learning or privacy-preserving analytics, which can share model improvements and leave the data local. This is becoming actively deployed in cross-border networks, sensor deployment in the private sector, and other community sensing programs. These, however, make reproducibility difficult, demand strong security and authentication, and may cause a performance skew in case data distributions vary strongly among sites[54].

The other dimension of governance that influences scientific credibility and transfer of operations is openness. Frameworks wherein there are open-code systems and documented datasets, and they hold pipelines, are less difficult to check and extend, whereas closed frameworks can function well, yet ensuring they are verifiable by the community is challenging. This dimension has an impact on the strength of the conclusion that can be made and the transferability of the reported results in a review context.

3.8. Framework Archetypes Emerging from the Taxonomy

The five axes define an uninterrupted space, but again and again, the literature and practice are grouped into familiar archetypes that exhibit goals, limitations, and common engineering motifs. One of the archetypes revolves around edge-based QC and smart ingestion, in which preprocessing and validation are enabled by low latency and bandwidth awareness. The other archetype is cloud-native fusion and productization, which focuses on integrating, scaling, and delivering reliably via Application Programming Interfaces (APIs). A third archetype includes DA-amazon observation improvement, in which learned operators, prejudice correction, and unforeseen estimation are created to advantage assimilation value. The fourth archetype is the use of systems, which are DA-integrated and purport an influence through cycling experiments and operational-like assessment. One of the fifth archetypes is in the domain of digital twins and adaptive observing, whose element is the central role played by simulation and closed-loop decision-making[55].

These archetypes are not exclusive to each other, and hybrid systems can lie between archetypes, especially when edge components are combined with cloud fusion and assimilation impact evaluation. However, the archetypes are practical since they suggest varying maturity standards. As an example, an edge QC framework will be evaluated based on robustness, fail-safe behavior, and traceable QC semantics, whereas a framework with a DA-integrated one will be evaluated based on forecast impact, cycle stability, and observational space uncertainty[56].

3.9. Mapping the Taxonomy onto Table 1

The relative usefulness of **Table 1** is based on the ability of the readers to quickly discern the purpose of the framework, where it is used, the degree of its connection, self-sufficiency, management, and proof of validation. It is due to this reason that the taxonomy must be operationalized, set in the form of table fields which are specific to maneuvering the five axes and important descriptors. **Table 1** in an SCI-standard review is not supposed to be a list of methods, but it should allow comparison, which would show patterns like which layers of deployment dominate which tasks, which parts of uncertainty are explicitly modelled, how often the impact of forecasts is demonstrated, and which are the assumptions of governance that are common. **Table 1** summarizes the comparative taxonomy of the representative AI-driven structures reviewed in this paper[3].

Table 1. Taxonomy and comparison of representative AI-driven frameworks for next-generation atmospheric observation networks.

ID	Framework/Study	Observation Modalities	Target Variables/Products	Core Functions	Deployment Topology	Learning Paradigm + Update Mode	Coupling to Physics/DA	Autonomy	Uncertainty Representation	Evaluation Type	Openness & Readiness Notes
1	Automated quality assessment of citizen weather stations	Citizen/IoT weather stations + reference network	QC weights/filtered temperature observations	QC, sensor reliability weighting, outlier handling	Cloud/centralized	Machine learning (ML)/statistical learning; offline (case-study tuned)	Standalone analytics	Passive	Implicit via weighting; not full probabilistic UQ	Offline + regional case study	Research prototype; demonstrates automated QC for Volunteered Geographic Information (VGI)/citizen networks
2	Adopting citizen observations in operational weather prediction	Citizen weather stations	Assimilation-ready observations (after QC)	QC + integration into operational workflows	Hybrid (network + central processing)	Mixed (QC + screening); operational integration focus	DA-aware (operational adoption emphasis)	Passive	Typically via QC/filters and DA error models (study-dependent)	Operational context + assimilation-oriented discussion	Emphasizes practical adoption issues and operational constraints for citizen obs
3	ML approach to the observation operator for satellite radiance DA)	Satellite microwave brightness temperatures	Observation operator (model→radiance equivalent) for DA	Learned observation operator/fast surrogate	HPC/DA environment	Supervised ML; offline training	DA-aware to DA-integrated (observation-operator role)	Passive	Can provide modeled-equivalent BT; uncertainty handling varies by implementation	DA-oriented evaluation (paper focuses on observing operator role)	Targets a key DA integration point (observation operator)
4	ML-based observation operator for FY-4B GIIRS radiance DA	FY-4B GIIRS (infrared sounding radiances)	End-to-end ML radiative transfer/operator (clear-sky)	Learned RTM/observation operator for DA	HPC/DA environment	Supervised ML; offline training	DA-aware (OO for radiance assimilation)	Passive	Operator-focused; uncertainty treatment depends on DA use	DA-oriented/method validation	Explicitly framed as an ML observation operator for radiance assimilation
5	NOAA neural-network synthetic radar reflectivity estimator using GOES-R observations	GOES-R satellite observations	Synthetic radar reflectivity (proxy product)	Retrieval/cross-sensor translation	Cloud/HPC (central processing typical)	Supervised NN; offline training	Standalone analytics (can be DA-aware if adapted)	Passive	Often deterministic product; UQ not always primary	Offline + case-based validation	Practical for filling radar gaps or enhancing situational awareness
6	ML-assisted city-scale air temperature mapping from sparse observations	Sparse in situ + auxiliary urban inputs (study-dependent)	High-resolution urban temperature fields	Fusion/downscaling/gap-filling	Cloud/centralized	Supervised ML; offline training	Standalone analytics	Passive	Varies; commonly point estimates with optional uncertainty proxies	Offline + urban case evaluation	Illustrates “framework as productization” for high-res fields from sparse networks
7	ML-driven real-time air-quality assessment framework	Fixed + mobile AQ sensors + met inputs + satellite	Real-time AQ assessment + health-risk prediction	Multi-source fusion, real-time prediction services	Cloud/hybrid (multi-source ingestion)	ML models; operational/real-time framing	Standalone analytics (DA coupling not central)	Passive	Varies; may include risk probabilities	Real-time framework demonstration	Explicitly integrates multi-source streams into a real-time assessment framework
8	NVIDIA Earth-2 (AI-accelerated climate digital twin platform)	Multi-source (platform-level: models + data services)	Digital twin applications for weather/climate simulation & visualization	Digital twin platform, AI + physics workflows, services	Cloud/HPC platform	Supports AI training/inference + simulation workflows	Digital twin/simulation-coupled	Typically passive (can support scenario/what-if)	Platform-dependent (can support ensembles/UQ via workflows)	Platform/engineering validation; application-specific	Proprietary platform with reference services for climate digital twins
9	Destination Earth (DestinE) initiative and Data Lake	EU-scale digital twin ecosystem + data lake	Digital twins (extremes, climate adaptation) + data services	Digital twin + data lake + service platform	Cloud/HPC ecosystem	Workflow-driven; supports AI/analytics integration	Digital twin/simulation-coupled	Typically passive (interactive/on-demand workflows)	Includes UQ as part of the digital twin aims (implementation-dependent)	Program-level operationalization (release milestones)	Public initiative with core service platform + data lake operated by EUMETSAT
10	OGC SensorThings API (interoperable IoT observation framework)	Heterogeneous IoT sensing systems	Standardized observations + metadata access	Interoperability layer (data plane), tasking support	Edge↔Cloud enabling standard	Not an AI method; enables AI-ready ingestion/governance	Standalone infrastructure (enabler for DA/AI pipelines)	Supports “Tasking” conceptually (control interface), not AI	N/A (standard defines data interfaces)	Standard/spec validation via implementations	Open, royalty-free standard for unified IoT observations/metadata exchange

Table 1. *Cont.*

ID	Framework/Study	Observation Modalities	Target Variables/Products	Core Functions	Deployment Topology	Learning Paradigm + Update Mode	Coupling to Physics/DA	Autonomy	Uncertainty Representation	Evaluation Type	Openness & Readiness Notes
11	Pangeo community + Pangeo-ML (scalable open geoscience + ML workflows)	Earth-system data (incl. atmospheric archives)	Scalable analysis + ML pipeline enablement	Cloud-scale data access, reproducible workflows	Cloud/HPC-friendly ecosystem	ML workflows/tooling; not a single model	Infrastructure enabler (supports many framework types)	Passive	Depends on the specific ML workflow	Workflow reproducibility focus	Open community + tooling to support scalable, reproducible geoscience ML
12	RL + DA framework for optimizing drone-based observations for flux estimation	Drone-based atmospheric observations	Improved surface flux estimation + sampling strategy	DA for estimation + RL for adaptive sampling	Hybrid (field platform + central compute)	RL policy + DA; trained offline with simulation/experiments	DA-aware + adaptive observing	Adaptive	Value via improved inference; UQ via DA/ensembles (study-dependent)	Method demonstration (framework-level)	Explicit closed-loop observing: DA inference + RL sampling policy

Table 1 is associated with a larger synthesis as well. The review may discover research gaps by clustering rows based on taxonomy bins, including the lack of focus on the calibrated uncertainty in edge deployments, the lack of assimilation-based validation of learned observation operators, or insufficient interoperable metadata/QC standards of federated networks.

3.10. Critical Comparison of Framework Performance and Failure Modes

Even though many AI-based observation systems are able to show encouraging potential in areas including data fusion, anomaly detection, and environmental search, there is a great degree of variance in the performance of these systems when applied in different settings. The quality of training data and consistency of an observational regime are critical to reported improvements. Learning-based models can experience degradation in predictive reliability under distributional shifts, such as extreme weather conditions or sensor failures[57].

There are a number of common limitations that have been reported in studies. To begin with, retrieval models that are data-driven can spread biases that were present in the training data or calibration errors. Second, multi-modal fusion architectures can enhance sensor artifacts in case quality-control processes are inadequate. Third, close integration of AI modules and data assimilation systems may lead to instability in the event of poor uncertainty estimates. These remarks reveal the necessity of strong assessment procedures and working protective measures in the implementation of AI elements into monitoring pipelines in the atmosphere[58,59].

This section specifies a framework-based taxonomy of the organization of AI-based systems in atmospheric observation networks. The taxonomy, in terms of introducing a framework as end-to-end, deployable systems as well as categorizing them on five main axes (deployment topology, learning and update mode, physical model, assimilation coupling, autonomy level, and governance model), gives a more consistent way of interpreting the comparative table and synthesizing the literature. Taxonomy also establishes a relationship between system type and appropriate evaluation standards and general risks, enabling subsequent sections to describe architectures, coupling strategies, and readiness considerations in a descriptive, operationally grounded manner[60].

4. End-to-End Architectures and Core Capabilities

The concept of AI-based atmospheric observation systems should be viewed as end-to-end designs that convert heterogeneous sensor measurements into an actionable product through validation and uncertainty-aware output, usable for scientific analysis, operational forecasting, or real-time decision support. Even though individual implementations differ greatly, most frameworks can be divided into a series of functional layers that are repeated in both modalities and deployment settings. This section provides an overview of these general architectural designs and the fundamental functionality they enable, with particular emphasis on how design decisions are influenced by latency, bandwidth, reliability, and the need to preserve provenance and uncertainty throughout the processing chain. This is not aimed at defining a single best architecture, but rather at determining some reference patterns according to which systems in **Table 1** can be compared based on their operationalization of intelligence, handling constraints, and environmental trustworthiness through the gaming lifecycle of an observing network[61].

4.1. Reference Architecture for AI-Enabled Observation Networks

On a higher level, next-generation observation frameworks apply a pipeline that will start with the acquisition of data and will culminate in products or services that will be consumed by human beings, models, or automated control logic. Sensing and local preprocessing, transmission (ingestion), quality control and calibration, harmonization and generation of features, fusion and inference, characterization of uncertainty, and presentation lot (delivery) are components of the pipeline, which utilize standard interfaces. Coupled with numerical models, the pipeline, in addition to observation-space transformations and assimilation-compatible outputs, will also have latency and reproducibility conditions[62]. To enhance better conceptual accessibility, the architectural structure that is presented in the present section is sketched schematically in **Figure 1**, which depicts the distinction between the data plane, which performs the responsibility of acquiring, processing, and inferencing the observations, and the control plane that takes charge of orchestration, monitoring, and model lifecycle management[63]. This abstraction can be used to explain the way that heterogeneous infrastructures of sensing can be incorporated into scalable AI-based systems of observation.

Another conceptual differentiation that is useful is between the control plane and data plane. Observations are carried and transformed in the data plane, and the control plane controls the system on how to run and how to manage configuration, model updates, monitoring, alerting, access control, and incident response. These planes are addressed in mature structures. This difference is significant because most operational failures are not due to model error but rather to control-plane failures, including the failure to configure, thus aging seamlessly, inexplicable drift, or the unavailability of rollback mechanisms to succeed an unsuccessful update. The control plane underlies scientific plausibility and operational reliability in networks of observation, which are more dependent on streaming and continuous model upkeep and enhancement[64]. **Table 2** gives an overview of the end-to-end pipeline functional decomposition, the likely inputs/outputs, and provenance requirements.

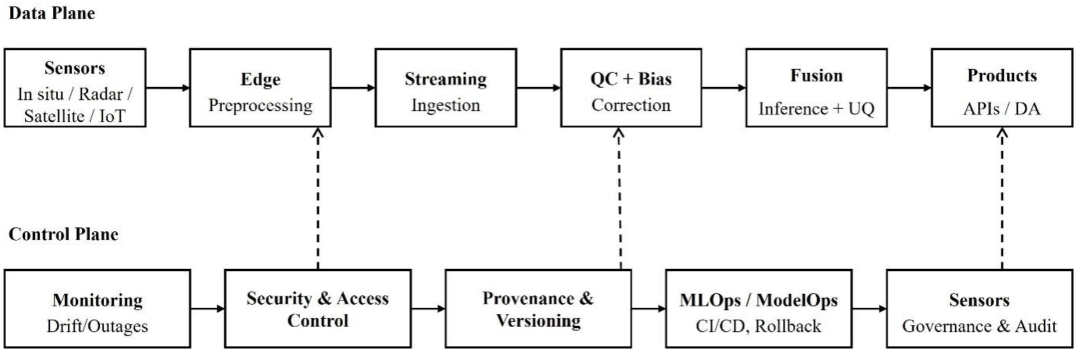

Figure 1. End-to-end reference architecture of AI-enabled atmospheric observation networks.

Table 2. Reference end-to-end architecture for AI-enabled atmospheric observation networks.

Architectural Layer	Typical Inputs	Processing/AI Roles	Primary Outputs	Key Constraints	Required Provenance/Metadata
Sensing & Edge Preprocessing	raw instrument signals, device diagnostics	on-device QC, anomaly/event detection, compression	validated packets, sensor health states	power, compute, intermittent connectivity	device ID, calibration state, model/firmware version
Transmission & Ingestion	streams/batches, mixed cadence	schema validation, deduplication, time alignment	trusted ingested observations	throughput, ordering, time drift	timestamps, ingestion logs, schema version

Table 2. *Cont.*

Architectural Layer	Typical Inputs	Processing/AI Roles	Primary Outputs	Key Constraints	Required Provenance/Metadata
QC & Calibration	ingested observations + neighbors/background	drift correction, QC flags, bias estimation	corrected values + QC semantics	false accept/reject cost	QC flag ontology, correction provenance
Harmonization & Feature Generation	multi-source data	unit/coord/vertical harmonization, representations	aligned features for fusion/DA	scale mismatch, missingness	coordinate reference, feature definitions
Fusion/Inference	multi-modal features	spatiotemporal fusion, gap filling, nowcasting	gridded fields, event probabilities	latency, scaling, robustness	uncertainty metadata, lineage links
Delivery & Services	products & logs	APIs, files, alerting, access control	decision-ready products	SLA, uptime, backward compatibility	product versioning, audit trail
Lifecycle (MLOps/ModelOps)	metrics, drift signals, verification	monitoring, CI/CD, rollback, governance	safe updates + diagnostics	security, reproducibility	dataset hashes, model registry IDs

4.2. Sensing, On-Device Preprocessing, and Intelligent Ingestion

The physical layer next to physical sensors in the digital pipeline is the first architectural layer. In the past, sensors generated measurements that were sent out mostly in their pure form, and quality control was done centrally. Various contemporary designs, in contrast, have on-device preprocessing, which involves basic diagnostics, timestamp validation, format normalization, and early-stage QC. The reason behind this change is the amount of data generated in dense networks and high-resolution remote sensing products, and operational limitations in remote or bandwidth-constrained environments.

Edge intelligence supports a number of functions that are becoming more fundamental. The former is fast anomaly detection and self-diagnostics to detect faulty sensors before they contaminate end products. The second one is drift and calibration check, especially with the sensors of low cost or in harsh conditions, atmospheric bias can drift faster than the maintenance. The third is smart compression and summarization, according to which the system only sends the information required by downstream purposes, which may be needed in case of sudden falls in pressure, appearance of convective signs, or spikes in pollutants. These functions minimize the bandwidth, as well as the latency, and enhance the performance of data completeness.

But going all the way to the extreme with intelligence brings its requirements in the design. The heterogeneity of hardware and the lack of resources needed to implement the model should be supported by model deployment, with inference optimization and sound quantization policies needed to guarantee deterministic behavior at resource constraints. Mechanisms of updates should be secure, authenticated, and reversible. Besides, edge outputs should preserve semantic clarity. When the edge device produces QC flags or corrections, the results should be standardized and have metadata for the model version, thresholds, and confidence measures at which these results were produced. In its absence, no scientific interpretation and auditing is possible later, and assimilation systems can misuse edge-modified observations[65].

4.3. Data Transport, Streaming Ingestion, and Metadata/Provenance Management

After the digital pipeline is fed with the observations, the ingestion layer has to maintain high-throughput streaming data and delayed or batch transfers. Architectures are becoming more based on the message-oriented design whereby various constituents are able to use the same stream of data to serve various functions, including near-real-time monitoring, product generation, and archival. Consumption is not only a throughput issue, but is also where the initial point of control over data can be intended by the use of schema control, authentication, and access control measures.

At this stage, metadata and provenance management come into the picture. These observations that are made in the atmosphere can only be interpreted when this is coupled with instrument identifiers, calibration parameters, geolocation, and altitude reference, sampling cadence, and processing history. The addition of provenance requirements by AI-enabled structures: model identifiers, training data, inference configuration, and uncertainty calibration metadata. One of the most important architectural trends is to consider provenance not as a secondary artefact but as a primary one. This consists of records of immutable event logs, versioned schemas, and systems to identify each derived product with the precise upstream inputs and model versions on which it was created. These are necessary to allow reproducibility and to detect failure in extremes, where misjudgment may be costly[66].

Time is another big issue. Most products of observation are latency-sensitive, and architectures need to be able to handle out-of-order arrival, nonlinear transmission delay, and clock skew of devices. Contemporary architectures commonly adopt event-time processing principles in which data are now reconciled by observation time instead of arrival time, and this allows strong transfer of intermittent connectivity. This becomes specifically important in combination with mobile sensor streams with fixed networks or remote sensing products, where timing discrepancies might instead manifest themselves as bodily errors[67].

4.4. Quality Control, Bias Correction, and Harmonization Modules

The fundamental effectiveness that secures downstream inference, fusion, and assimilation is quality management. When architecturally, QC frequently is not one phase, but it consists of fast and conservative checks at the early stages of the pipeline, as well as/more context-sensitive checks at later stages. The invalid records that are obviously invalid can be eliminated during the initial QC phase, and sensor restrictions can be applied; subsequent phases might use spatial and temporal context, or previous background model fields, or allow shear materials discussion to mark possible inconsistencies. In AI-based frameworks, QC modules could consist of learned anomaly detectors, probabilistic consistency checking, or graph-based, which can use correlations across dense networks[68].

QC is commonly used with bias correction and drift management. Systematic errors are the worst in most networks. Frameworks thus have the advantage of having explicit modules that approximate the components of bias, their development, and corrections. The architectural issue is that the bias correction is to be considered as the uncertainty modeling and provenance, and not as some obscure transformation. The downstream consumers, particularly assimilation systems, should be aware of whether an observation has been fixed or not, and the way its uncertainty should be understood after correction.

Multi-modal frameworks are also dependent on harmonization. It involves the regular coordinate reference frames, unit normalization, vertical references, and translation of heterogeneous measurements into similar ones. The harmonization is also applied to the semantic interoperability, e.g., uniformity between sensor types and locations in QC flags. Models that do not harmonize tend to yield products that are ostensibly plausible, but which degrade as more sensors are introduced, or the network is stretched over a wider geographical area, or when some other institution is trying to combine the results[69].

4.5. Feature Generation, Retrieval, and Representation Learning

In addition to QC and harmonization, most frameworks contain a layer that converts observations to higher-level representations. This can be in remote sensing, where geophysical variables can be retrieved in terms of radiances or reflectivities. In high-density in situ networks, it can be the building of spatial features, temporal aggregates, or neighborhood structure representations. In pipelines with AI, the representation learning is a unifying process that has the capacity to extract features that can be used in multiple tasks, including gap-filling and fusion tasks, as well as nowcasting.

Physics determines the design of this layer. Often, retrieval tasks are ill-posed, and several atmospheric conditions are consistent with the same measurement. Consequently, there is a tendency towards frameworks being represented in a probabilistic manner or generating distributions and confidence intervals, and not just a point estimate. On the one hand, when physical constraints are known (i.e., not influenced by the underlying object being represented), they may be represented in the model architecture or implicitly as loss functions and post-processing constraints. In both scenarios, the framework has to maintain traceability in order to ensure that the users can tell what was measured, what was inferred, and what assumptions were made during the retrieval process[70].

It is more important in convenient samples of labels that are seldom known or faulty. Self-supervised methods can use huge archives to discover embeddings that capture spatiotemporal structure, sensor cross-correlations, and regime signatures. Such embeddings may be transferred to local networks and jobs, which may enhance resistance to noisy data and variability. Nevertheless, the advantage of representation learning relies on the prudent assessment in the face of distribution shift and means of observing the instances when learned representations become inconsistent with the operational reality[71].

4.6. Multi-Modal Fusion and Spatiotemporal Modeling

Next-generation observation structures are characterized by fusion. Fusion may take place on multiple levels, both architecturally, or on the raw-data level, at the feature-level, or at the decision-level. The strategy selected is based on the volumes of data, the quality of alignment, and the necessity to be interpretable and propagate uncertainty, e.g., radar and satellite products can be fused at the feature level following harmonization, and in situ networks can be fused with graph-based models that model sensors as a graph with learned or physically informed relationships[72,73].

The fusion of atmospheric fields is particularly based on the spatial-temporal modeling due to the structure and multiscale dynamics of atmospheric fields. Irregular sampling, irregular cadence, and missingness are also models that are commonly included within frameworks. One of the architectural choices that keeps recurring is whether to do early imposition of grid-based representations to be able to be compatible with downstream modeling systems, or do delayed gridding by using irregular representations retaining local detail. Early gridding is easy to implement but exposes one to the problem of smoothing extremes, banning, and obfuscation of representativeness errors; late gridding conserves resolution but also makes product generation and assessment more complex[74].

The uncertainties propagation in the light of fusion can be considered as one of the key variables of the distinction between research prototypes and operationally credible frameworks. In case fusion outputs should be used to support decisions or assimilation, they should bear the measures of uncertainty that indicate the error in observation, uncertainty in models, and the extent of agreement between modalities. When uncertainty is designed as an ad hoc phenomenon, architectures tend to create overconfident fields that meet conditions seldom, which is when trust is most needed, and it is this that compromises trust[75].

4.7. Productization, Interfaces, and Service Delivery

The end map of the framework is productization: transforming processed observations into deliverables, which can be both consumed on a delicious and repeatable basis. To scientific users, it can also imply versioned datasets including rich metadata, coordinated QC semantics, and a record of processing histories. To the operational users, it can imply low-latency gridded fields, alerts, or probabilistic hazard guidance, which is available with the aid of APIs. In the case of modeling systems, this can be in the form of assimilation-ready observation files, estimates of bias, and observation-error specifications consistent with the DA system into which it is consumed[76].

One of the fundamental architectural principles is that even though the internal models of a product may change, the product interfaces must remain stable. This demands deep consideration of schema formation and backward compatibility. It also demands explicit product versioning and the capability of recreating past production, which is critical to scientific audit, performance regression analysis, and troubleshooting post-occurrence of significant events. A large number of frameworks thus implement service-oriented designs in which data products are modeled as fixed releases, and metadata that identifies the sources of data, processing stages, and versions of the AI models[77].

4.8. Lifecycle Engineering: MLOps/ModelOps for Observation Frameworks

The AI-driven systems of observation are not fixed. Sensors become off course, networks grow, algorithms become better, and the requirements of users change. Lifecycle engineering is hence included in the core architecture. Organizational practices associated with MLOps/ModelOps are dataset versioning, model registries, automated testing, continuous evaluation, pipelines of deployment, drift and degradation monitoring, and drift and degradation incident response in this case.

The networks of observation present an MLOps issue of particular difficulty since loops of feedback are either feeble or slow. Ground truth can be unrealized on a real-time basis, and thus, there can be difficulty in quickly detecting a performance decay. The frameworks commonly use proxy measures like consistency with local sensors, consistency of error distributions, or consistency with fields of the background model. In the high-stakes applications, it is directed to have both the statistical signals as well as the domain-based alarms, along with the escalation process and fallback avenue. Governance is also essential in the hybrid and edge deployments, especially when it comes to updates. Thousands of devices will be impacted by a model update difference, and a spatially localized artifact can be produced by non-uniform rollout. Due to this, frameworks enjoy the advantages of staged deployments, canary tests, and the capability to roll back to older versions of the models. Auditability means that all outputs can be linked to the precise model and configuration that was employed, which implies controlled control-plane design [78–80].

4.9. Cross-Layer Design Considerations: Trustworthiness, Robustness, and Interoperability

There are some design considerations that cut across all the layers of the architecture. The attributes of being trustworthy require maintaining the provenance, clearly characterizing uncertainty, and the ability to audit and be able to reproduce. Robustness is based on the ability to deal with distribution shift, missingness, sensor failures, and regime switches, and robustness to malicious inputs in exposed networks. Interoperability requires standard metadata, semantic consistency in QC, and stable interfaces so that products are shared throughout the institutions and ingested by the downstream systems without any special integration effort [81,82].

Such requirements interact. As an example, a framework that works on aggressive edge correction should minimize random noise but also disrupt provenance unless it records that it has corrected the image transparently. Generalization can be enhanced with a framework that is based on a large foundation model, but makes reproducibility difficult when training data and model weights are not stored and versioned. Optimizing the latency framework can complexity uncertainty calculation accomplishments in the event that lightweight stochastic frameworks are devised in-architecture, and can be more successful in the transition between proof-of-concept demonstrations to scalable infrastructures.

This section has generalized the end-to-end architectural designs that cross-cut across AI-based atmospheric observation systems. Starting with sensing and intelligent ingestion, transitioning to streaming ingestion and provenance management, and culminating with QC, retrieval, fusion, and product delivery, the architectures presented here are descriptions of how the frameworks operationalize AI within the real framework of the world. It was noted in the discussion that lifecycle engineering, lifecycle monitoring, and interoperability were not peripheral issues, but they are characteristic features of next-generation systems. On these architectural foundations, Section 5 studies the way structures couple to physical models and data assimilation and how they facilitate observing strategies adaptively, which makes a circular link between sensing and prediction [83].

5. Integration with Physical Models, Data Assimilation, and Adaptive Observing

AI-based observation systems can be maximized with the maximum achievable scientific and functional value when productively interacting with physical models. In atmospheric science, there are dynamical structure theatres (state models), and physically consistent state evolution (numerical weather prediction—NWP, chemical transport models, and coupled Earth-system models), along with data assimilation (data assimilation), furnishing the formal structures of the combination of observations and model prediction. The difficulty- and the opportunity- is that AI of these days could enhance many aspects of this coupled system; however, only when it is put in a manner that does not disrupt stability, meaning, and plausible uncertainty. In this segment, the three major routes to Bonaparte AI system linkage to physical frameworks and the conventional designs of assimilation-sensitive outputs, as well as the substance in the new direction of closed-loop/adaptive observation, where the sensing strategies are guided by model sensitivity and forecast uncertainty, are described [5,84].

5.1. Why Coupling Matters: From Stand-Alone Skill to Forecast and Analysis Impact

This has remained a consistent weakness in the field of AI-for-atmosphere studies, as the indicators employed to assess models are not always comparable to those employed to evaluate the status of an observation network. Stand-alone predictive performance, though helpful, does not tell one that a processed observation will enhance an analysis or a prediction[85]. The effects of an observation in DA are conditioned by the error statistics, representativeness of the observation with respect to the model state, its bias properties, the structure of spatial and temporal correlation, as well as the background error properties of the model. A component of an AI which alters the mean error of a retrieval but mis-calibrates uncertainty, such as can harm assimilation as an over-weight to mis-calibrated information. Equally, a hyper-aggressive AI-based QC system can also eliminate desirable extreme-event signals and decrease forecast skill in the areas where it is most needed. Due to this reason, assimilation-conscious integration must incorporate structures to provide more than corrected values. They have to provide uncertainty and provenance in forms that can be interpreted by the DA systems, and they should provide a clear document of assumptions and stable cyclic behavior. This puts it in a different way of focusing on the best model to best couple behavior, and it can clarify why integration is sometimes the most difficult stage in applying AI innovations to daily operations[86–88].

5.2. Assimilation-Oriented Outputs: Observation Operators, Errors, and Bias Handling

Assimilation systems are usually performed in the observation space by the use of an observation operator that projects a model state onto what an instrument would measure. Most of the most significant integration points of AI frameworks happen here. In remote sensing, a physical forward model or retrieval chain can be replaced or accelerated with a learned operator, allowing a fast way to work with nonlinearity or a better way to work with nonlinearity. Mapped representations of point measurements into grid-cell counterparts (representativeness mappings) are important in in situ networks where learned operators can be less central[89].

A paradigm that tries to justify DA must thus explain which of the three roles it would fulfill. The former is generating an estimation of an observation in model-consumable variables, which may be in the form of a retrieval with volatility. The second one is the generation of an observation operator or its surrogate, which allows a DA system to compare the model and observation in a mutual measurement space. The third one is generating meta-information that DA needs: estimates of observation error, bias corrections, quality flags, and correlation assumptions.

The one component of AI integration that can be the weakest is observation-error specification. DA systems depend on the relative weighting between the background and observation information. When the AI-based uncertainties are underestimated, the system will be over-confident and capable of enhancing the errors; when overestimated, the possible improvements will be dampened. Furthermore, spatial and temporal correlations of errors in observation are common because of common instruments, data processing, and the effects of the environment. These correlations are not explicitly modeled in frameworks, but can be dominant in assimilation behavior. One of the main design suggestions derived from the literature is that assimilation-based frameworks ought, at least, to give calibrated estimates of uncertainty in observation space, and clearly state whether correlations are neglected, estimated, or being modeled[90,91].

Discrimination management is also basic. Most DA systems make the assumption that they have unbiased errors or include bias correction schemes. AI models that conduct implicit bias elimination without reporting can be inconsistent with such schemes and cause either doubled correction or remaining biases. Frameworks that are aware of assimilation are thus advantaged in that they expose estimates of bias as distinct products, monitor their change over time, and maintain provenance such that configurations of the DA can be tuned in a similar way as a model or sensor evolves[89].

5.3. Hybrid Modeling and Physics-Informed Integration

One of the main reasons why AI should be combined with physical models is that atmospheric processes are subject to constraints that could be broken by pure data-driven learning, particularly when there is a change in distribution. Hybrid modeling attempts to balance the power of expressiveness of AI and the stability and understanding of physics. Hybrid methods take many forms in observation models: as learning with physical correction residuals, as physical monotonicity and bounds being regulated in the loss, as a physical conservation-like regularization, or as a physical architecture design that obeys known physical monotonicity and bounds[92].

Physics-Informed Deep Priors (PIDPs) are most important for ill-posed problems in retrieval as well as sparse training data regimes, necessitating rare extrema or poorly observed phenomena. Structures can facilitate generalization and decrease the likelihood of unrealistic results by pegging learned elements to a physically significant structure. Nevertheless, the physics-informed design does not imply correctness; the constraints can be incomplete or broken by the processes outside of the model. It is the readiness of operations to perform that relies, then, on how well hybrid systems are assessed with respect to average skill, but also constraint violations, regime errors, and extrapolatory behavior[93].

The second type of integration that is important is the application of AI surrogates to speed up the costly parts of physical models. Surrogate operators are able to lower the cost of computation, allowing fewer cycles or larger ensembles to be used to estimate uncertainty. However, surrogates should be established to be stable in the coupled loop. Even tiny errors of approximation can compound when under cycling, and stability can be determined by the operating regime. Those frameworks that involve the aid of the surrogates ought to thus record their area of validity and encompass safeguards that conceptualize when contributions are beyond taught regimes[94].

5.4. Coupling Pathways: From "DA-Aware" to "DA-Integrated"

Section 3 provides the taxonomy that differentiates between the frameworks that are DA-aware and the frameworks that are DA-integrated. In reality, the coupling process can be in between. A DA-aware system can first provide better QC and uncertainty estimates that are ingested by the existing assimilation systems but make no changes to the fundamental DA machinery. The burden of integration in this pathway is less, and it can be incrementally validated. In the long run, it is possible to have the work of integrating frameworks where learned operator functionalities, bias models, or adaptive observation error schemes have been directly incorporated into the assimilation workflow. **Table 3** outlines the main integration directions, including those of standalone analytics and those of DA-integrated and digital-twin systems.

Table 3. Integration pathways with physical models and data assimilation.

Pathway Class	What the AI Framework Contributes	What Must Be Output Explicitly	Typical Value Added	Common Failure Modes	Appropriate Validation Evidence
Standalone Analytics	QC/fusion/retrieval/nowcasting products	uncertainty + provenance + QC semantics	rapid utility, low integration burden	overconfidence, hidden bias, brittle preprocessing	offline tests + regime/extremes stratification
DA-Aware (External)	assimilation-ready obs products	obs-space errors, bias terms, representativeness notes	improves DA inputs with minimal DA changes	misweighted obs degrade DA	DA trials + innovation stats diagnostics
DA-Integrated	learned operator/correction inside the DA cycle	stable interfaces, version control, cycling-safe UQ	measurable analysis/forecast gains	cycling instability, error correlations ignored	OSE cycling across multiple cases/regimes
Digital Twin/OSSE	simulated network + measurement behavior	simulator fidelity + uncertainty	safe pre-deployment testing	sim–real mismatch	OSSE + field "reality checks"
Adaptive Observing	policies for sensing/transmission prioritization	action logs, constraints, override rules	information-efficient sensing	feedback loop bias, unsafe autonomy	shadow-mode + staged rollout + OSSE

This difference is important due to the fact that it influences the evidence required to support adoption. DA-aware systems may be tested using offline measures, as well as restricted assimilation experiments, but DA-integrated systems must exhibit robust cycling and prediction of impact in a variety of cases and regimes. Moreover, systems that contain DA also have to deal with operational guarantees, namely deterministic reproducibility, runtime budgets, and hard configuration control, which might be unknown in a standard ML research environment[95].

Another repeated finding amongst coupled systems is that integration is a co-design problem. The assimilation configuration cannot be optimized without the processing of observations. The decisions relating to bias correction, thinning, superbeing, and correlation assumption and QC thresholds interact very well with the outputs of the AI module. Frameworks that transparently model these dependencies and have tunable interfaces stand a better chance of success than those based on a one-way, improved observation and assimilation will improve the pipeline[96,97].

The progression from standalone processing to closed-loop adaptive observing, and the corresponding increase in validation and governance requirements, is illustrated in **Figure 2**.

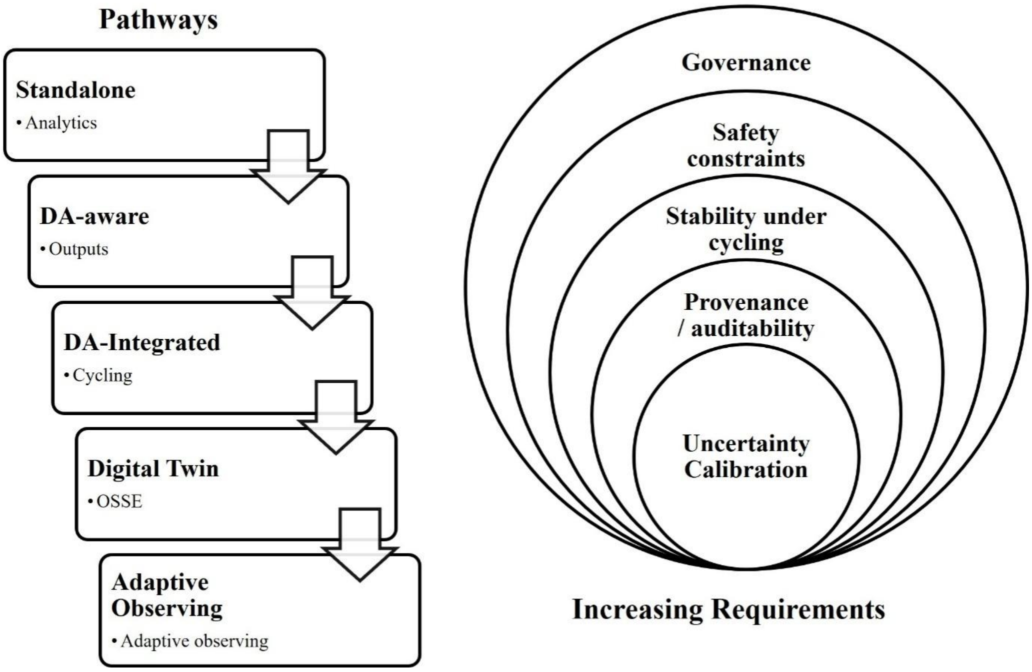

Figure 2. Coupling pathways: From standalone analytics to DA-integrated and digital-twin systems.

5.5. Observing System Experiments and Impact Attribution

In situations where frameworks say they can make predictions better, the most important methodological issue is the problem of attribution: how can one separate the impact of the AI-enabled observation processing from other concomitant shifts in the modeling system? Observation and systems verification. All the standard tools are observation system experiments and observation system simulation experiments. OSEs analyze the effect of actual observations by comparing the analysis and forecasts with and without a particular observation type or processing method. OSSEs are based on a simulated truth atmosphere and simulated observations that can be used to test new sensors or processing strategies in advance[98].

The frameworks based on AI raise new complexities in OSE/OSSE design. Considering that a framework involves online learning, the with and without experiment results can separate over time as a result of the varying data streams employed to update them. Where the framework also produces uncertainty estimates, any variation in assimilation weighting could be a significant effect, and can only be effectively diagnosed. Provided the framework alters QC behavior, it can be case-dependent and advantageous to certain regimes but damaging to others. Strict review of the SCI-standard, therefore, highlights that forecast impact claims ought to be buttressed by numerous cases, stratification of regimes, as well as diagnostics that distinguish between the effects of mean bias and the effects of variance and uncertainty. Where feasible, the frameworks must report not only aggregate skill scores, but also analysis increments, statistics on innovation, and measures of reliability that will tell whether the assimilation system is performing as expected with errors[98].

5.6. Digital Twins of Observation Networks

Digital twins do not limit coupled models of the atmosphere to explicit models of the observation network itself. A network digital twin may contain sensor locations and scheduling, communication constraints, failure and maintenance models, and approximate physics of measurement. Digital twins, in combination with an atmospheric simulator or with archived conditions, provide a means to systematically test their sensor placement, make them redundant, or adaptive, without placing them in a real-world situation[99].

Digital twins have three significant purposes in the framework of AI-driven schemes. They supply artificial information to be coached and attempt, particularly in unique extremes and stress conditions that can be prevented in historical archives. They make it possible to evaluate new framework components in an OSSE-like fashion, e.g., learned QC thresholds or adaptive sampling policies. They also promote lifecycle choices, including giving maintenance or calibration choices in which information worth is most critical.

Fidelity is critical to the processes of generating believable conclusions about digital twin results. Assuming that the twin is not a realistic model picture of communication delays, sensor drift, and retrieval artifacts, an adaptive strategy might be helpful in simulation but not work in action. Consequently, the digital twin frameworks are recommended to focus on the uncertainty of the simulator itself and are expected to be tested against the actual network behavior. It is a common idea: coupled cyber-physical systems are also uncertain about the state of the atmosphere, measurement models, and infrastructure models[100].

5.7. Adaptive Observing and Closed-Loop Frameworks

Adaptive observing structures are the closest manifestation of the convergence of codes and climate: the observation system becomes an active participant in the predictive system and distributes sensing resources in accordance with the forecasted information benefits. These frameworks can work at various levels. They might determine the positioning of new sensors at the network level, or they can arrange the mobile platforms. On the operational level, they can cause increased sampling frequency, give priority to transmissions, or steer platforms to areas of good forecast sensitivity. At the model level, they can select the observations to assimilate or the thinning and superb according to estimated error correlations.

Information theory and decision theory are often the starting point of the theoretical foundation of adaptive observing. Strategies can be designed to maximize the reduction in forecast uncertainty that is likely to occur, to minimize the target variable expected loss, or to maximize mutual information between candidate observations and quantities of interest. Practically, high-dimensional systems cannot be reached by full value-of-information computations, as such computations are computationally intensive, and approximations are needed. Reinforcement learning, bandit, and active learning AI techniques can deliver a set of tractable choice policies, and ensembles and adjoint sensitivity techniques can deliver approximations of forecast uncertainty and sensitivity[101].

Adaptive frameworks raise the levels of safety and governance. The errors are amplified when they operate through feedback loops since they change what data are being collected or transmitted. As an illustration, a policy that is always focused on well-monitored areas is prone to overlooking underserved areas, diminishing equity, and failing to detect emerging risks. A policy that is excessively optimistic about the estimates of uncertainty can seize upon spurious signals. Furthermore, unscrupulous control of the edge devices or communication channels may divert the sensing resources in malicious directions. Thus, closed-loop structures demand clear restrictions, fail-safe systems, human control, and logs of decisions. In most contexts, the most realistic containing near-term solution is neither full autonomy nor distinguishing somewhat (destined), although familiar with human-in-loop reactive viewing, in which AI suggests, and operators authorize or modify them[102].

Operational observation networks make use of human-in-the-loop decision-making devices less and less to schedule sensors, investigate anomalies, or switch to an adaptive sampling scheme more often. Judgment by humans is also required on safety-critical actions, especially in situations where there is extreme weather surveillance or mission- crucial satellites. Under such environments, AI is used more as a decision-support layer, with operational control being left to the expert operators. The idea of a hybrid human-AI process is thus a significant step between the existing state of a set of more-or-less isolated observational networks and the future state of a completely autonomous adaptive observing system[103].

5.8. Integration Challenges: Consistency, Stability, and Trust in Coupled Systems

The inclusion of AI in DA and adapting to observing fractures presents a list of challenges that are qualitatively unlike those that are seen in independent tasks with ML. Consistency is one of the first concerns: how the processing of observations, the estimates of uncertainties, and bias corrections are consistent with the assumptions made in the assimilation system has to be. Stability is also very essential: learned parts should not cause oscillatory behavior or drift and corruption of performance with cycling, and online updates have to be strictly controlled. Trust is based upon traceability, reproducibility, and the capability to diagnose failures in a coupled loop where errors have the possibility to cascade through a number of components.

These hurdles bring into focus the need to employ frameworks as opposed to models as the unit of analysis. An excellent performance in offline retrieval does not imply the assimilation readiness of a retrieval model. A targeting policy that seems successful on a simulated platform is not invariably safe on a working platform. Combining AI with physical models thus requires a rigorous practice due to the integration that needs to take place and encompass domain theory, rigorous evaluation, and strong software practices. In this section, the way AI-based observation models can be coupled with physical models via assimilation-oriented outputs, hybrid and physics-informed designs, and increasing levels of closeness between DA-aware pipelines and DA-integrated systems is reviewed. It highlighted that rational ambiguity, openness of parties to prejudice, and robustness to cyclic updating are requirements of forecast-impact claims. It was then coupled to network digital twins and adaptive observing and emphasized the potential improvements in information efficiency as well as the increased demands on safety, governance, and auditability of closed-loop systems. These integration lines pre-empt the next step of integrating, Section 6, which is an integration of evaluation practices, operational readiness requirements, cross-cutting challenges, and future directions to establish next-generation free, interoperable, authentic infrastructures of atmospheric observation[104].

6. Evaluation, Readiness, Cross-Cutting Challenges, and Future Directions

Evaluation and comparison of AI-based frameworks of atmospheric observation networks are multi-dimensional in nature due to the variety of such frameworks. A structure can be scientifically strong but operationally weak or effective in computation but badly calibrated, or work in average cases but fail in sequences[105]. Furthermore, the final usefulness of an observation scheme is often indirectly achieved by way of better analyses, predictions, and decision performance, but not by model-only scores. This part, then, condenses practice evaluation suited to the various classes of framework, suggests a readiness-based view that is in line with operational anticipations, and summarizes cross-cutting issues that continue to narrow translation among research prototypes and persistent infrastructure. Based on these premises, it subsequently gives future directions that could be used to design next-generation, trustworthy, and interoperable observing systems[106].

6.1. Evaluation as a Multi-Layer Problem

Observation framework assessment should take into account the stratification of the pipeline between observation and product. Each AI part can be assessed at the lowest level using classical ML metrics, although these metrics are only indicative of some of what is relevant in atmospheric applications. At the system level, the framework should be evaluated in terms of reliability, latency, and robustness in the case of nonstationary as well as failure. A relevant question at the impact level is, does the framework ameliorate downstream scientific or operating consequences, such as the quality of assimilation, forecasting skill, and risk-relevant measurement of decisions?

These layers are coupled. An even minor change in a measure of retrieval error can have no effect on forecasts when uncertainties are poorly specified, or representativeness prevails. On the other hand, observations that are only moderately accurate can also be much more useful in forecasts, provided that the error in their observations is fully characterized and they have available information in sensitive areas. It is then necessary to design an evaluation to fit the role and level of coupling that the framework is intended to play as specified by the taxonomy in Section 3[107].

6.2. Component-Level Evaluation: Skill, Robustness, and Interpretability

On the component level, assessment starts with predictive skill on held-out data, though to have a credible assessment of atmospheric observation, it is not sufficient to have aggregate measures of accuracy. The failure modes that are prevalent usually include being biased during a certain regime, being sensitive to sensor artifacts, and being degraded under distribution shift. Based on this, stratification of performance should be done based on meteorological regime, season, geography, sensor type, and extreme event conditions. In the case of QC modules, the evaluation should be based on the asymmetry of the costs: rejecting good data results in a false negative, whereas acceptance of corrupted data will pollute the products of fusing and assimilation[108].

The robustness assessment must clearly cover patterns of missingness, intermittent data streams, timing jitter, and simulated outage. Observation structures are often run on incomplete information; none of the performance assertions that assume that there is clean and full information are operationally equivalent. Also applicable at this level, interpretability is important, especially in the context of QC and bias correction, in which scientific users and operational operators should be aware of why data are flagged or fixed. It is not intended to give full transparency of model internals, but to have actionable interpretability with meaningful categories of QC, measures of confidence, and diagnostics that are auditable[109].

Direct qualitative comparison between AI-based observation systems is not an easy task because of the variation in datasets, evaluation procedures, and operation settings. Rather than trying to do a possibly misleading meta-analysis, this review suggests a stratified evaluation system that differentiates component-level skill, system-level operational performance, and the impact of the downstream forecast. Such a hierarchy is based on the established practices in the research on atmospheric data assimilation, and it offers a systematic foundation for future empirical benchmarking.

6.3. Uncertainty Quantification and Calibration Evaluation

The most significant product of an observation framework is typically uncertainty, particularly when the output is used to make assimilation or risk-sensitive choices. Accuracy evaluation should thus be given equal attention as calibration evaluation. The uncertainty outputs of a framework ought to be experimented with respect to reliability, such as the predicted intervals or probabilities sensitive to actually observed frequencies across regimes, as well as in the presence of a shift. Perceptiveness, without stability, is of no service to it; stable but too conservative uncertainty may also be destructive of value, by under-weighting the observations.

Frameworks that traverse edge and cloud layers should also be able to guarantee the consistency of definitions of uncertainty. As an example, an edge QC model may give a confidence score, whereas a cloud fusion model may give probabilistic fields, and uncertainty is not coherently modeled and may be double-counted or even ignored in downstream tasks. It should thus be evaluated not only to check the calibration of each of the modules but also to check the consistency of the uncertainty propagation across the pipeline. In the case of uncertainty used to weight observations to DA, the analysis of innovation statistics and other diagnostic statistics should be used to determine whether the assumed errors are similar to the actual discrepancies between model background and observations[110].

6.4. System-Level Evaluation: Latency, Reliability, and Operational Performance

System-level evaluation determines the capability of the framework to provide its desired products under the constraints of real life. Latencies are dependent on the application, and could be seconds to minutes in case of hazard nowcasting, to hours or days in climatic monitoring. Frameworks have to be tested against the declared latency budgets at realistic loads, peak data, and partial failure conditions. Operational credibility revolves around metrics of reliability, including uptime, recovery time, growth in backlog, and subpar connectivity and behavior during component failures.

The evaluation equation should cover the cost and resources (bandwidth consumption, edge power draw) in more resources where applicable, as well as training and inference compute footprint. Such considerations do not just comprise engineering specifications; they define the ability of a framework to scale, whether it is sustainable, or even whether it will be possible to implement it in areas with limited resources. Energy and carbon factors become more important along with model and network size, and these are more applicable to 24/7 continuous inference pipelines. System-level evaluation should also include security and resilience. Spoofing and tampering, as well as supply-chain risks, are prone to distributed observation networks. Frameworks possessing adaptive decision-making or edge intelligence should be assessed in terms of being resilient to corrupted inputs as well as safe when the assumptions of trust are not followed[111].

6.5. Impact-Level Evaluation: Assimilation and Forecast Skill

Impact-level evaluation is necessary for frameworks that are supposed to support DA or make predictions better. The conventional procedure includes monitoring system experimentation and associated assimilation tests, comparing and predicting with and without the processing or products of the framework. There should be impact evaluation across various cases and regimes, the diagnostics of which identify improvements because of bias reduction, variance reduction, improved weighting of uncertainty, or improved representativeness.

The right environment to investigate the possibility of probabilistic benefits to provide improved reliability of forecast products is also an impact evaluation. The outcomes of interest in the case of decision support can be enhanced warning lead time, a smaller number of false alarms, or more accurate probabilities relating to stakeholder-interest thresholds. Due to the sensitivity of forecasts and decisions to extremes, we ought to give direct consideration to the rare-event performance and must not be guided by global averages that will conceal regime-specific harms[112,113].

The evaluation hierarchy used in this review from component skill to downstream impact, is depicted in **Figure 3.**

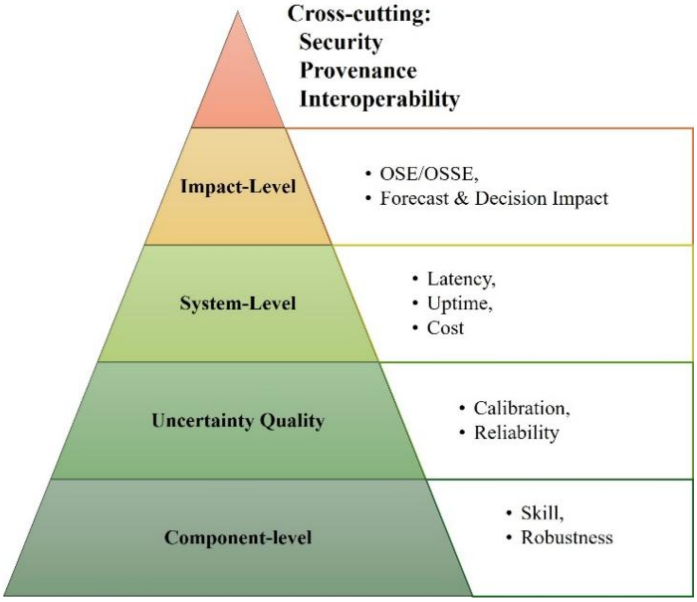

Figure 3. Evaluation hierarchy and metrics stack.

6.6. Operational Readiness and Reproducibility

Operational preparedness must have a framework that can be maintained, audited, and replicated. Reproducibility is a scientific virtue as well as an empirical requirement when we are trying to diagnose failure or do post-event analyses. A readiness-oriented assessment thus encompasses the versioning of datasets and preprocessing pipelines, whether models are versioned, whether the outputs can be traced to inputs and model versions as well, and the system will also support rollback in an update.

Overseeing is an element of preparedness. Due to the frequent unavailability of ground truth in real time, monitoring needs to integrate a mix of both statistical messages, like drift indicators and distribution checks, and domain-sensitive diagnostics, including spatial consistency measures, innovation statistics relative to background fields, and sensor health measures. Human factors also constitute preparedness: signifying well-defined alerting routes, documentation, and interfaces that would allow the operators to grasp and act upon anomalies. **Table 4** summarizes the minimum evidence suggested to be credible to the operations.

Ready Interoperability is included in readiness. Abstracting Software Like Frameworks should have identifiable interfaces and follow metadata and QC semantics and provenance conventions to allow downstream systems to absorb the outputs with custom engineering. In the case of multi-agency ecosystems, governance agreements regarding updates, responsibility regarding incident response, and acceptable use of the shared model are also considered to be part of readiness[114,115].

Table 4. Evaluation and operational readiness checklist for AI-driven observation frameworks.

Dimension	What to Report	Examples of Measures	Minimum "SCI-Review Credible" Reporting Standard
Component Skill	accuracy and error decomposition	RMSE/MAE, bias, detection metrics	include regime + extremes breakdown
Uncertainty Quality	calibration and reliability	coverage, CRPS, reliability diagnostics	show calibration under shift
Robustness	behavior under failure/shift	outage tests, missingness tests, drift indicators	include stress testing protocol
System Performance	latency/reliability/cost	end-to-end latency, uptime, bandwidth/power	measured under realistic load
Security/Resilience	threat awareness + mitigations	authentication, tamper detection, audit logs	documented control plane + incident plan
DA/Forecast Impact	coupled-system improvement	OSE/OSSE, skill deltas, innovation stats	multi-case cycling evidence
Reproducibility	traceability	dataset/model versioning, config capture	output → inputs/models traceability

6.7. Benchmarking and Standardized Evaluation Protocols

Standardized benchmarking strategies are needed in a progressive direction towards the operational adoption of AI-enabled observation networks. A good benchmark suite must offer a wide range of assessment cases, i.e., extreme weather regimes, sensor failure cases, multi-modal data fusion problems, and experiments with simulated observation systems. These standards ought to measure predictive performance and other operational characteristics such as robustness, latency, and reliability. Setting up common evaluation policies will enhance consistency in research and facilitate the transformation of AI-based observation systems between research prototypes and working infrastructure.

6.8. Equity, Coverage, and Geographic Bias

The optimization of sensing networks with the help of AI can accidentally contribute to the reinforcement of geographic inequalities in the coverage of the observations. Areas with a high concentration of historical data can be overtrained and optimized in model training processes, whereas areas with little data are underrepresented. The solution to those difficulties would be to take clear steps in considering equity in the network design, such as using bias-aware sensor participation plans, federated learning models, and international data-sharing programs to enhance global observation coverage. **Figure 4** is a roadmap that brings together interoperability, unified lifecycle tooling, trustworthy UQ to extremes, federated learning governance and digital-twin maturity[116–119].

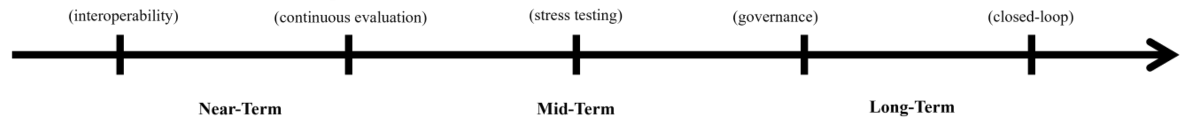

Figure 4. Research and deployment roadmap for next-generation observing frameworks.

7. Conclusions

AI transforms atmospheric observation networks not only through the refinement of specific algorithms, but through the possible allowance of a novel genre of end-to-end designs, considering perceiving and processing as a kettle of fish. Through the literature that has been reviewed in this paper, a recurring theme has been identified; the effectiveness and worth of next-generation observing systems are increasingly becoming software architecture, lifecycle engineering, and governance decisions rather than sensor hardware or discrete model expertise. In this regard, convergence of code and climate can be comprehended as moving away from the assumption of data collection and offline analysis to continuously running, uncertain, interoperable infrastructures that can learn, adapt, and provide actionable products at scale.

This review gave a framework-based synthesis of the discipline. To enable heterogeneous systems to be evaluated on a shared footing, we proposed a taxonomy based on five characteristic axes of AI-driven observation frameworks, including deployment topology, learning and update mode, connection to physical models and data assimilation, degree of autonomy, and governance model. It is based on this taxonomy that we summarized repeating architectural designs across edge intelligence, streaming ingestion, quality control and bias management, multi-modal fusion and product delivery, highlighting the fact that provenance, interoperability, and ModelOps/MLOps are fundamental scientific demands as opposed to optional engineering extensions. Integration pathways in physical models and data assimilation were also discussed, showing that forecast-impact claims require there to be calibrated uncertainty, explicit bias control, and cycling stability, and that digital twins and adaptive observing open a path to closed-loop sensing but cost a lot of safety and accountability. Lastly, we have merged assessment and preparedness deliberations, contending that reliable validation should not only be evaluated in terms of predictive accuracy but also in terms of nonstationary and extremes, system reliability and latency, security and resilience, reproducibility, and provable downstream effect.

Each of the framework classes has several cuts across gaps and priorities. First, uncertainty is the currency of trustful integration: lack of coherent and calibrated uncertainty flowing through QC, fusion and assimilation even very good models can distort coupled performance and cause loss of trust by the user. Second, interoperability/provenance has been a consistent bottleneck in an augmented observing system—the bankrolled AI-ready semantics. A standardized semantics is now required to support QC, bias corrections and uncertainty plus lineage auditing operations can support scientific reproducibility and operational accountability. Third, the distribution shift and the extremes are also to be designed as the first-order design constraints instead of the corner cases since it is the observing system which is of the greatest concern when the conditions are not the norm of the past. Fourth, frameworks that are more distributed and adaptive are also characterized by security, existing coverage marginality, and institutional control as determining whether the benefits may be achieved at a large scale and not subject to continuous encouragement. Fifth, system design should be made sustainable and energy efficient, particularly with the growing use of continuous inference and training of large models.

Though AI-based observation schemes have significant potential in enhancing the ability to monitor atmospheric environments, future studies are needed to go beyond the idea frameworks to provide empirical evidence and make them operational. The top priorities are the creation of assimilation-based AI evaluation metrics, the strong uncertainty propagation through distributed pipelines, and energy-efficient AI architectures, which can be used to address large-scale sensing networks. Also, the interoperable metadata standards, as well as digital-twin simulation environments, will be the key to the environment of safe and auditable adaptive observing systems. To move these directions forward, the cooperation of atmospheric scientists, machine learning scientists, and operational forecasting organizations will be very close.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created during this work.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] National Research Council, 2009. Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks. National Academies Press: Washington, DC, USA. DOI: https://doi.org/10.17226/12540

[2] Bluestein, H.B., Carr, F.H., Goodman, S.J., 2022. Atmospheric Observations of Weather and Climate. Atmosphere-Ocean. 60(3–4), 149–187. DOI: https://doi.org/10.1080/07055900.2022.2082369

[3] Zaidan, M.A., et al., 2025. Artificial Intelligence for Atmospheric Sciences: A Research Roadmap. arXiv preprint. arXiv:2506.16281. DOI: https://doi.org/10.48550/arXiv.2506.16281

[4] Davenport, T., Harris, J., 2017. Competing on Analytics: Updated, with a New Introduction: The New Science of Winning. Harvard Business Press: Boston, MA, USA.

[5] Zhao, T., Wang, S., Ouyang, C., et al., 2024. Artificial intelligence for geoscience: Progress, challenges, and perspectives. The Innovation. 5(5), 100691.

[6] Shi, W., Zhang, M., Zhang, R., et al., 2020. Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges. Remote Sensing. 12(10), 1688. DOI: https://doi.org/10.3390/rs12101688

[7] Kumar, G.K., Sankar, P., 2025. Artificial Intelligence Powered Satellite Communications and Sentinel Satellite Constellations: An Overview and Future Perspectives. SN Computer Science. 6(6), 735. DOI: https://doi.org/10.1007/s42979-025-04268-8

[8] Singh, M., Arora, V., Kulshreshta, K., 2024. AI and the Environment: Innovative Approaches to Climate Change. In: Singh, B., Kaunert, C., Vig, K., et al. (Eds.). Practice, Progress, and Proficiency in Sustainability. IGI Global: Hershey, PA, USA. pp. 1–22. DOI: https://doi.org/10.4018/979-8-3693-6336-2.ch001

[9] Gheibi, O., Weyns, D., 2022. Lifelong self-adaptation: Self-adaptation meets lifelong machine learning. In Proceedings of the 17th Symposium on Software Engineering for Adaptive and Self-Managing Systems, Pittsburgh, PA, USA, 18–23 May 2022; pp. 1–12. DOI: https://doi.org/10.1145/3524844.3528052

[10] Max, L.B. Pereira, A., 2024. The IC AI Multiplier: Automating Superiority—Seizing Adversarial Artificial Intelligence use in Intelligence Operations. DOI: https://doi.org/10.2139/ssrn.4884351

[11] Bourechak, A., Zedadra, O., Kouahla, M.N., et al., 2023. At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives. Sensors. 23(3), 1639. DOI: https://doi.org/10.3390/s23031639

[12] Wang, F., Zhang, M., Wang, X., et al., 2020. Deep Learning for Edge Computing Applications: A State-of-the-Art Survey. IEEE Access. 8, 58322–58336. DOI: https://doi.org/10.1109/ACCESS.2020.2982411

[13] Wang, H., Guo, R., Ma, P., et al., 2025. Towards Mobile Sensing with Event Cameras on High-agility Resource-constrained Devices: A Survey. arXiv preprint. arXiv:2503.22943.

[14] Lynch, J.P., Wang, Y., Swartz, R.A., et al., 2008. Implementation of a closed-loop structural control system using wireless sensor networks. Structural Control and Health Monitoring. 15(4), 518–539. DOI: https://doi.org/10.1002/stc.214

[15] Choi, Y., Yoon, S., 2023. In-situ observation virtual sensor in building systems toward virtual sensing-enabled digital twins. Energy and Buildings. 281, 112766. DOI: https://doi.org/10.1016/j.enbuild.2022.112766

[16] Kamran, M., Tanveer, K., Khalid, N., et al., 2025. Integrating Advanced Deep Learning Algorithms For Climate Systems: Enhancing Weather Forecast Accuracy, Real-Time Climate Monitoring, And Long-Term Climate Predictions. Spectrum of Engineering Sciences. 3(6), 365–403.

[17] Bandi, P.K.R., 2025. The Role of Metadata in Making Data AI-Ready: Enhancing Data Discoverability and Usability. Journal of Computer Science and Technology Studies. 7(5), 954–963.

[18] Verhulst, S., Zahuranec, A., Chafetz, H., 2025. Moving Toward the FAIR-R principles: Advancing AI-Ready Data. DOI: https://doi.org/10.2139/ssrn.5164337

[19] Baño-Medina, J., Sengupta, A., Michaelis, A., et al., 2025. Harnessing AI Data-Driven Global Weather Models for Climate Attribution: An Analysis of the 2017 Oroville Dam Extreme Atmospheric River. Artificial Intelligence for the Earth Systems. 4(3), 240090. DOI: https://doi.org/10.1175/AIES-D-24-0090.1

[20] Widergren, S.E., Narang, D., Khandekar, A., et al., 2018. Interoperability Strategic Vision. Pacific Northwest National Laboratory (PNNL): Richland, WA, USA.

[21] David, I., Shao, G., Gomes, C., et al., 2025. Interoperability of Digital Twins: Challenges, Success Factors, and Future Research Directions. In: Margaria, T., Steffen, B. (Eds.). Leveraging Applications of Formal Methods, Verification and Validation. Application Areas, Lecture Notes in Computer Science.

Springer Nature: Cham, Switzerland. pp. 27–46. DOI: https://doi.org/10.1007/978-3-031-75390-9_3

[22] Marinescu, M.-R., Ionescu, O.N., Pachiu, C.I., et al., 2025. Next-Gen Healthcare Devices: Evolution of MEMS and BioMEMS in the Era of the Internet of Bodies for Personalized Medicine. Micromachines. 16(10), 1182. DOI: https://doi.org/10.3390/mi16101182

[23] Schnase, J.L., Lee, T.J., Mattmann, C.A., et al., 2016. Big Data Challenges in Climate Science: Improving the next-generation cyberinfrastructure. IEEE Geoscience and Remote Sensing Magazine. 4(3), 10–22. DOI: https://doi.org/10.1109/MGRS.2015.2514192

[24] Shen, Z., Jin, J., Tan, C., et al., 2024. A Survey of Next-generation Computing Technologies in Space-air-ground Integrated Networks. ACM Computing Surveys. 56(1), 1–40. DOI: https://doi.org/10.1145/3606018

[25] Mois, G., Sanislav, T., Folea, S.C., 2016. A Cyber-Physical System for Environmental Monitoring. IEEE Transactions on Instrumentation and Measurement. 65(6), 1463–1471. DOI: https://doi.org/10.1109/TIM.2016.2526669

[26] Ahmed, S.H., Kim, G., Kim, D., 2013. Cyber Physical System: Architecture, applications and research challenges. In Proceedings of the 2013 IFIP Wireless Days (WD), Valencia, Spain, 13–15 November 2013; pp. 1–5. DOI: https://doi.org/10.1109/WD.2013.6686528

[27] Rawat, D.B., Rodrigues, J.J.P.C., Stojmenovic, I. (Eds.), 2015. Signal and Data Processing Techniques for Industrial Cyber-Physical Systems. In Cyber-Physical Systems. CRC Press: Boca Raton, FL, USA. pp. 200–245. DOI: https://doi.org/10.1201/b19290-16

[28] Aires, F., Aznay, O., Prigent, C., et al., 2012. Synergistic multi-wavelength remote sensing versus a posteriori combination of retrieved products: Application for the retrieval of atmospheric profiles using MetOp-A. Journal of Geophysical Research: Atmospheres. 117(D18), 2011JD017188. DOI: https://doi.org/10.1029/2011JD017188

[29] Monteleone, S., Alves De Moraes, E., Protil, R.M., et al., 2024. Proposal of a Model of Irrigation Operations Management for Exploring the Factors That Can Affect the Adoption of Precision Agriculture in the Context of Agriculture 4.0. Agriculture. 14(1), 134. DOI: https://doi.org/10.3390/agriculture14010134

[30] Zhou, Q., Su, X., Wang, A., et al., 2013. QC-Chain: Fast and Holistic Quality Control Method for Next-Generation Sequencing Data. PLoS ONE. 8(4), e60234. DOI: https://doi.org/10.1371/journal.pone.0060234

[31] Tan, R., Xing, G., Liu, B., et al., 2012. Exploiting Data Fusion to Improve the Coverage of Wireless Sensor Networks. IEEE/ACM Transactions on Networking. 20(2), 450–462. DOI: https://doi.org/10.1109/TNET.2011.2164620

[32] Goodess, C.M., Hall, J., Best, M., et al., 2007. Climate Scenarios and Decision Making under Uncertainty. Built Environment. 33(1), 10–30. DOI: https://doi.org/10.2148/benv.33.1.10

[33] Murphy, A., Katz, R.W., 2019. Probability, Statistics, and Decision Making in The Atmospheric Sciences. CRC Press: Boca Raton, FL, USA. DOI: https://doi.org/10.1201/9780429303081

[34] Von Clarmann, T., Degenstein, D.A., Livesey, N.J., et al., 2020. Overview: Estimating and reporting uncertainties in remotely sensed atmospheric composition and temperature. Atmospheric Measurement Techniques. 13(8), 4393–4436. DOI: https://doi.org/10.5194/amt-13-4393-2020

[35] Hüllermeier, E., Waegeman, W., 2021. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine Learning. 110(3), 457–506. DOI: https://doi.org/10.1007/s10994-021-05946-3

[36] Safaeipour, F.Z., Chakareski, J., Hashemi, M., 2025. Bayes-Split-Edge: Bayesian Optimization for Constrained Collaborative Inference in Wireless Edge Systems. In Proceedings of the Tenth ACM/IEEE Symposium on Edge Computing, Arlington, VA, USA, 3–6 December 2025; pp. 1–13. DOI: https://doi.org/10.1145/3769102.3770629

[37] Bocquet, M., Elbern, H., Eskes, H., et al., 2015. Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. Atmospheric Chemistry and Physics. 15(10), 5325–5358. DOI: https://doi.org/10.5194/acp-15-5325-2015

[38] Zhang, S., Liu, Z., Zhang, X., et al., 2020. Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: A review. Climate Dynamics. 54(11–12), 5127–5144. DOI: https://doi.org/10.1007/s00382-020-05275-6

[39] Tamburri, D.A., Lago, P., Vliet, H.V., 2013. Organizational social structures for software engineering. ACM Computing Surveys. 46(1), 1–35. DOI: https://doi.org/10.1145/2522968.2522971

[40] Myers, C.R., 2003. Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs. Physical Review E. 68(4), 046116. DOI: https://doi.org/10.1103/PhysRevE.68.046116

[41] Nickerson, R.C., Varshney, U., Muntermann, J., 2013. A method for taxonomy development and its application in information systems. European Journal of Information Systems. 22(3), 336–359. DOI: https://doi.org/10.1057/ejis.2012.26

[42] Ghil, M., 1989. Meteorological data assimilation for

oceanographers. Part I: Description and theoretical framework. Dynamics of Atmospheres and Oceans. 13(3–4), 171–218. DOI: https://doi.org/10.1016/0377-0265(89)90040-7

[43] Zhou, B., Zhang, S., Xue, R., et al., 2023. A review of Space-Air-Ground integrated remote sensing techniques for atmospheric monitoring. Journal of Environmental Sciences. 123, 3–14. DOI: https://doi.org/10.1016/j.jes.2021.12.008

[44] Guo, Y., Huo, P., Huang, S., et al., 2025. Multi-Receptor Skin with Highly Sensitive Tele-Perception Somatosensory Flexible Electronics in Healthcare: Multimodal Sensing and AI-Powered Diagnostics. Advanced Healthcare Materials. 14(25), 2502901. DOI: https://doi.org/10.1002/adhm.202502901

[45] Diaz, M., 2011. Forming a Definitional Framework for "Intelligence". American Intelligence Journal. 29(1), 53–64.

[46] Tumushime-Mugisha, A.G., 2020. Development and Validation of an Early-Stage Project Evaluation Tool [Master's Thesis]. Malardalen University: Västerås, Sweden.

[47] Gupta, A., Faraboschi, P., Gioachin, F., et al., 2016. Evaluating and Improving the Performance and Scheduling of HPC Applications in Cloud. IEEE Transactions on Cloud Computing. 4(3), 307–321. DOI: https://doi.org/10.1109/TCC.2014.2339858

[48] Zhang, X., Ming, X., Liu, Z., et al., 2019. A reference framework and overall planning of industrial artificial intelligence (I-AI) for new application scenarios. The International Journal of Advanced Manufacturing Technology. 101(9–12), 2367–2389. DOI: https://doi.org/10.1007/s00170-018-3106-3

[49] Burström, T., Parida, V., Lahti, T., et al., 2021. AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research. Journal of Business Research. 127, 85–95. DOI: https://doi.org/10.1016/j.jbusres.2021.01.016

[50] Camporese, M., Girotto, M., 2022. Recent advances and opportunities in data assimilation for physics-based hydrological modeling. Frontiers in Water. 4, 948832. DOI: https://doi.org/10.3389/frwa.2022.948832

[51] Choi, S.R., Lee, M., 2023. Transformer Architecture and Attention Mechanisms in Genome Data Analysis: A Comprehensive Review. Biology. 12(7), 1033. DOI: https://doi.org/10.3390/biology12071033

[52] Shmueli, E., Singh, V.K., Lepri, B., et al., 2014. Sensing, Understanding, and Shaping Social Behavior. IEEE Transactions on Computational Social Systems. 1(1), 22–34. DOI: https://doi.org/10.1109/TCSS.2014.2307438

[53] Benford, S., Schnädelbach, H., Koleva, B., et al., 2005. Expected, sensed, and desired: A framework for designing sensing-based interaction. ACM Transactions on Computer-Human Interaction. 12(1), 3–30. DOI: https://doi.org/10.1145/1057237.1057239

[54] Gomes, R.C., Liddle, J., Gomes, L.O.M., 2010. A Five-Sided Model Of Stakeholder Influence: A cross-national analysis of decision making in local government. Public Management Review. 12(5), 701–724. DOI: https://doi.org/10.1080/14719031003633979

[55] Manu, A., 2024. Moving Past Archetypes. In Transcending Imagination. Chapman and Hall/CRC: Boca Raton, FL, USA. pp. 165–184. DOI: https://doi.org/10.1201/9781003450139-17

[56] Balmer, J., Swisher, M., 2019. Diagramming the Big Idea: Methods for Architectural Composition. Routledge: London, UK.

[57] Lima, V.G.P., Andrade, E., Araújo, D.R.B.D., 2025. Dependability Analysis of Weather Monitoring Systems Considering Different Redundancy Mechanisms. In Proceedings of the 26th Symposium on High-Performance Computing Systems (SSCAD), Bonito, Brazil, 28–31 October 2025; pp. 13–24. DOI: https://doi.org/10.5753/sscad.2025.15308

[58] Alotaibi, E., Nassif, N., 2024. Artificial intelligence in environmental monitoring: In-depth analysis. Discover Artificial Intelligence. 4(1), 84. DOI: https://doi.org/10.1007/s44163-024-00198-1

[59] Shukla, K.K., Gupta, A., Arora, S.V., et al., 2025. Environmental monitoring and protection using artificial intelligence. In Proceedings of the 2nd International Conference on Optimization Techniques in Engineering & Technology, Greater Noida, India, 14–15 June 2024. DOI: https://doi.org/10.1063/5.0285889

[60] de Arcaya, J.D., 2024. A Framework for the Operationalization of Analytic Workloads in Complex Distributed Computing Environments [PhD Thesis]. Universidad de Deusto: Bilbao, Spain.

[61] Ali, H., Safdar, R., Liu, J., et al., 2025. Hybrid Fusion Paradigm in Advanced Process Monitoring: A Panoramic Review and Future Perspectives. Industrial & Engineering Chemistry Research. 64(47), 22465–22514. DOI: https://doi.org/10.1021/acs.iecr.5c02759

[62] Iqbal, M.W., Ashfaq, U., Soomro, A.A., et al., 2025. Towards Next-Generation Automation: Data-Driven Synergies of AI and Robotics through Data Engineering and Data Science. Spectrum of Engineering Sciences. 3(9), 181–209.

[63] Michel, O., Bifulco, R., Rétvári, G., et al., 2022. The Programmable Data Plane: Abstractions, Architectures, Algorithms, and Applications. ACM Computing Surveys. 54(4), 1–36. DOI: https://doi.org/10.1145/3447868

[64] Michailidis, E.T., Potirakis, S.M., Kanatas, A.G., 2020. AI-Inspired Non-Terrestrial Networks for IIoT: Review on Enabling Technologies and Applications.

IoT. 1(1), 21–48. DOI: https://doi.org/10.3390/iot1010003

[65] Esmail, M.A., Fathallah, H., 2013. Physical Layer Monitoring Techniques for TDM-Passive Optical Networks: A Survey. IEEE Communications Surveys & Tutorials. 15(2), 943–958. DOI: https://doi.org/10.1109/SURV.2012.060912.00057

[66] Murarka, S., Jain, A., Singh, L., 2024. Advanced Techniques in Data Ingestion and Pipelining for Scalable Big Data Platforms: A Comprehensive Review. In Proceedings of the 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, 13–14 December 2024; pp. 1–6. DOI: https://doi.org/10.1109/ICTBIG64922.2024.10911053

[67] Briscoe, B., Brunstrom, A., Petlund, A., et al., 2016. Reducing Internet Latency: A Survey of Techniques and Their Merits. IEEE Communications Surveys & Tutorials. 18(3), 2149–2196. DOI: https://doi.org/10.1109/COMST.2014.2375213

[68] Xiahou, X., Chen, G., Li, Z., et al., 2025. Knowledge Management in Construction Quality Management: Current State, Challenges, and Future Directions. IEEE Transactions on Engineering Management. 72, 1069–1088. DOI: https://doi.org/10.1109/TEM.2025.3550354

[69] Li, W., Peng, Y., Zhang, M., et al., 2023. Deep model fusion: A survey. arXiv preprint. arXiv:2309.15698. DOI: https://doi.org/10.48550/arXiv.2309.15698

[70] Papadimitroulas, P., Brocki, L., Christopher Chung, N., et al., 2021. Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization. Physica Medica. 83, 108–121. DOI: https://doi.org/10.1016/j.ejmp.2021.03.009

[71] Patel, J., Bolton, T.A., Schöttner, M., et al., 2025. Structural Connectome Harmonization Using Deep Learning: The Strength of Graph Neural Networks. arXiv preprint. arXiv:2507.13992. DOI: https://doi.org/10.48550/arXiv.2507.13992

[72] Li, Y., Xiao, X., 2025. Deep Learning-Based Fusion of Optical, Radar, and LiDAR Data for Advancing Land Monitoring. Sensors. 25(16), 4991. DOI: https://doi.org/10.3390/s25164991

[73] Snidaro, L., Visentini, I., Foresti, G.L., 2011. Data Fusion in Modern Surveillance. In: Remagnino, P., Monekosso, D.N., Jain, L.C. (Eds.). Innovations in Defence Support Systems–3, Studies in Computational Intelligence. Springer: Berlin, Heidelberg. pp. 1–21. DOI: https://doi.org/10.1007/978-3-642-18278-5_1

[74] Qian, H., Wang, M., Zhu, M., et al., 2025. A Review of Multi-Sensor Fusion in Autonomous Driving. Sensors. 25(19), 6033. DOI: https://doi.org/10.3390/s25196033

[75] Tsanousa, A., Bektsis, E., Kyriakopoulos, C., et al., 2022. A Review of Multisensor Data Fusion Solutions in Smart Manufacturing: Systems and Trends. Sensors. 22(5), 1734. DOI: https://doi.org/10.3390/s22051734

[76] Palo, J., 2023. Product Roadmapping Tool and Process Unification as a Part of Global End-to-End Repeatability Operating Model Development [Master's Thesis]. Haaga-Helia University of Applied Sciences: Helsinki, Finland.

[77] Bourahli, A., 2020. Approaches to Productization Processes: A Case by Junttan Oy [Bachelor's Thesis]. Savonia University of Applied Sciences: Kuopio, Finland.

[78] Chen, L., Xia, C., Zhao, Z., et al., 2024. AI-Driven Sensing Technology: Review. Sensors. 24(10), 2958. DOI: https://doi.org/10.3390/s24102958

[79] Sharma, A., Sharma, V., Jaiswal, M., et al., 2022. Recent Trends in AI-Based Intelligent Sensing. Electronics. 11(10), 1661. DOI: https://doi.org/10.3390/electronics11101661

[80] George, A.S., 2024. Leveraging IoE and AI for Continuous Observation of Human Social Dynamics. Partners Universal Innovative Research Publication. 2(5), 1–17.

[81] Ali, N., Baker, S., O'Crowley, R., et al., 2018. Architecture consistency: State of the practice, challenges and requirements. Empirical Software Engineering. 23(1), 224–258. DOI: https://doi.org/10.1007/s10664-017-9515-3

[82] Appelbaum, D., 2016. Securing Big Data Provenance for Auditors: The Big Data Provenance Black Box as Reliable Evidence. Journal of Emerging Technologies in Accounting. 13(1), 17–36. DOI: https://doi.org/10.2308/jeta-51473

[83] M, A., M, D., Amaithi Rajan, A., et al., 2024. EdgeShield: Attack resistant secure and privacy-aware remote sensing image retrieval system for military and geological applications using edge computing. Earth Science Informatics. 17(3), 2275–2302. DOI: https://doi.org/10.1007/s12145-024-01256-z

[84] Xu, Y., Liu, X., Cao, X., et al., 2021. Artificial intelligence: A powerful paradigm for scientific research. The Innovation. 2(4), 100179.

[85] Freire, P.K.D.M.M., Santos, C.A.G., Silva, G.B.L.D., 2019. Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting. Applied Soft Computing. 80, 494–505. DOI: https://doi.org/10.1016/j.asoc.2019.04.024

[86] Lee, J.-Y., Wang, B., Kang, I.-S., et al., 2010. How are seasonal prediction skills related to models' performance on mean state and annual cycle? Climate Dynamics. 35(2–3), 267–283. DOI: https://doi.org/10.1007/s00382-010-0857-4

[87] Clarke, D.P., Al-Abdeli, Y.M., Kothapalli, G., 2013. The impact of renewable energy intermittency on

the operational characteristics of a stand-alone hydrogen generation system with on-site water production. International Journal of Hydrogen Energy. 38(28), 12253–12265. DOI: https://doi.org/10.1016/j.ijhydene.2013.07.031

[88] Penny, S.G., Hamill, T.M., 2017. Coupled Data Assimilation for Integrated Earth System Analysis and Prediction. Bulletin of the American Meteorological Society. 98(7), ES169–ES172.

[89] Yuan, X., Niu, G., Yin, J., et al., 2025. Data Assimilation in Hydrological Models: Methods, Challenges and Emerging Trends. Hydrology. 12(12), 323. DOI: https://doi.org/10.3390/hydrology12120323

[90] Hanna, N., Trzcina, E., Möller, G., et al., 2019. Assimilation of GNSS tomography products into the Weather Research and Forecasting model using radio occultation data assimilation operator. Atmospheric Measurement Techniques. 12(9), 4829–4848. DOI: https://doi.org/10.5194/amt-12-4829-2019

[91] Bayat, M., Alizadeh, H., Mojaradi, B., 2023. Assimilation versus optimization for SWAT calibration: Accuracy, uncertainty, and computational burden analysis. Water Supply. 23(3), 1189–1207. DOI: https://doi.org/10.2166/ws.2023.055

[92] Reichstein, M., Camps-Valls, G., Stevens, B., et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature. 566(7743), 195–204. DOI: https://doi.org/10.1038/s41586-019-0912-1

[93] Schönlieb, C.-B., Shumaylov, Z., 2025. Data-driven approaches to inverse problems. arXiv preprint. arXiv:2506.11732. DOI: https://doi.org/10.48550/arXiv.2506.11732

[94] Šturek, D., Lazarova-Molnar, S., 2025. Surrogate Modeling: Review and Opportunities for Expert Knowledge Integration. Procedia Computer Science. 257, 826–833. DOI: https://doi.org/10.1016/j.procs.2025.03.106

[95] Alrobai, A., McAlaney, J., Dogan, H., et al., 2016. Exploring the Requirements and Design of Persuasive Intervention Technology to Combat Digital Addiction. In: Bogdan, C., Gulliksen, J., Sauer, S., et al. (Eds.). Human-Centered and Error-Resilient Systems Development, Lecture Notes in Computer Science. Springer International Publishing: Cham, Switzerland. pp. 130–150. DOI: https://doi.org/10.1007/978-3-319-44902-9_9

[96] Chang, H., Zhang, D., 2019. Identification of physical processes via combined data-driven and data-assimilation methods. Journal of Computational Physics. 393, 337–350. DOI: https://doi.org/10.1016/j.jcp.2019.05.008

[97] Li, X., Wang, Z., Huang, Y., et al., 2023. A Survey on Self-Evolving Autonomous Driving: A Perspective on Data Closed-Loop Technology. IEEE Transactions on Intelligent Vehicles. 8(11), 4613–4631. DOI: https://doi.org/10.1109/TIV.2023.3319689

[98] Bach, J., Holzäpfel, M., Otten, S., et al., 2017. Reactive-Replay Approach for Verification and Validation of Closed-Loop Control Systems in Early Development. In Proceedings of the WCX$^{\mathrm{TM}}$ 17: SAE World Congress Experience, Detroit, MI, USA, 4–6 April 2017. DOI: https://doi.org/10.4271/2017-01-1671

[99] Sajadieh, S.M.M., Noh, S.D., 2025. From Simulation to Autonomy: Reviews of the Integration of Artificial Intelligence and Digital Twins. International Journal of Precision Engineering and Manufacturing-Green Technology. 12(5), 1597–1628. DOI: https://doi.org/10.1007/s40684-025-00750-z

[100] Samuel, P., Saini, A., Poongodi, T., et al., 2023. Artificial intelligence–driven digital twins in Industry 4.0. In Digital Twin for Smart Manufacturing. Elsevier: Amsterdam, The Netherlands. pp. 59–88. DOI: https://doi.org/10.1016/B978-0-323-99205-3.00002-X

[101] Morss, R.E., Emanuel, K.A., Snyder, C., 2001. Idealized Adaptive Observation Strategies for Improving Numerical Weather Prediction. Journal of the Atmospheric Sciences. 58(2), 210–232. DOI: https://doi.org/10.1175/1520-0469(2001)058%253C0210:IAOSFI%253E2.0.CO;2

[102] Trenberth, K.E., Anthes, R.A., Belward, A., et al., 2013. Challenges of a Sustained Climate Observing System. In: Asrar, G.R., Hurrell, J.W. (Eds.), Climate Science for Serving Society. Springer: Dordrecht, The Netherlands. pp. 13–50. DOI: https://doi.org/10.1007/978-94-007-6692-1_2

[103] Papenkordt, J., 2025. Teaming with AI: Human Behavior in AI-Assisted Decision-Making [PhD Thesis]. University of Paderborn: Paderborn, Germany.

[104] Kutbi, M., 2024. Artificial Intelligence-Based Applications for Bone Fracture Detection Using Medical Images: A Systematic Review. Diagnostics. 14(17), 1879. DOI: https://doi.org/10.3390/diagnostics14171879

[105] Mouradian, W.E., Huebner, C.E., 2007. Future Directions in Leadership Training of MCH Professionals: Cross-Cutting MCH Leadership Competencies. Maternal and Child Health Journal. 11(3), 211–218. DOI: https://doi.org/10.1007/s10995-006-0170-3

[106] Snow, C., 2007. Cross-cutting themes and future research directions. In Developing Reading and Writing in Second-Language Learners. Routledge: London, UK. pp. 289–314.

[107] Levitan, B.S., Andrews, E.B., Gilsenan, A., et al., 2011. Application of the BRAT Framework to Case Studies: Observations and Insights. Clinical Pharmacology & Therapeutics. 89(2), 217–224. DOI: https://doi.org/10.1038/clpt.2010.280

[108] You, R., Xu, F., Liu, M., 2025. SAR Aircraft Segmen-

tation With SAR-to-Optical Image Translation and Segment Anything Model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 18, 22082–22093. DOI: https://doi.org/10.1109/JSTARS.2025.3602288

[109] Chen, X., Singh, M.M., Geyer, P., 2024. Utilizing domain knowledge: Robust machine learning for building energy performance prediction with small, inconsistent datasets. Knowledge-Based Systems. 294, 111774. DOI: https://doi.org/10.1016/j.knosys.2024.111774

[110] Cho, H.-N., Choi, H.-H., Kim, Y.-B., 2002. A risk assessment methodology for incorporating uncertainties using fuzzy concepts. Reliability Engineering & System Safety. 78(2), 173–183. DOI: https://doi.org/10.1016/S0951-8320(02)00158-8

[111] Nouri, A., 2015. Rigorous System-Level Modeling and Performance Evaluation for Embedded System Design [PhD Thesis]. Université Grenoble Alpes: Grenoble, France.

[112] Sureau, S., Mazijn, B., Garrido, S.R., et al., 2018. Social life-cycle assessment frameworks: A review of criteria and indicators proposed to assess social and socioeconomic impacts. The International Journal of Life Cycle Assessment. 23(4), 904–920. DOI: https://doi.org/10.1007/s11367-017-1336-5

[113] Douthwaite, B., Kuby, T., Van De Fliert, E., et al., 2003. Impact pathway evaluation: An approach for achieving and attributing impact in complex systems. Agricultural Systems. 78(2), 243–265. DOI: https://doi.org/10.1016/S0308-521X(03)00128-8

[114] Johnny, R., Leory, B., 2024. Maintenance and Audit Readiness in Post-Disaster Building Management. Available from: https://www.researchgate.net/publication/386424749_MAINTENANCE_AND_AUDIT_READINESS_IN_POST-DISASTER_BUILDING_MANAGEMENT (cited 10 December 2025).

[115] Browne, S., Pike, T., Bailey, M.M., 2024. A proposed framework for artificial intelligence safety and technology readiness assessments for national security applications. OSF Preprints. DOI: https://doi.org/10.31219/osf.io/ekth8

[116] Petrillo, M., Fabbri, M., Kagkli, D.M., et al., 2022. A roadmap for the generation of benchmarking resources for antimicrobial resistance detection using next generation sequencing. F1000Research. 10, 80. DOI: https://doi.org/10.12688/f1000research.39214.2

[117] Malhotra, S., 2025. Next-Generation Observability Platforms: Redefining Debugging and Monitoring at Scale. SSRN Electronic Journal. DOI: https://doi.org/10.2139/ssrn.5190462

[118] Steinman, J.S., 2013. The Roadmap. In Proceedings of the 2013 Spring Simulation Interoperability Workshop, San Diego, CA, USA, 8–12 April 2013.

[119] Laflamme, S., Ubertini, F., Di Matteo, A., et al., 2023. Roadmap on measurement technologies for next generation structural health monitoring systems. Measurement Science and Technology. 34(9), 093001. DOI: https://doi.org/10.1088/1361-6501/acd135