International Astronomy and Astrophysics Research Journal

Volume 8, Issue 1, Page 13-48, 2026; Article no.IAARJ.162019

Physics-Informed Machine Learning in Astronomy and Astrophysics: A Methodology-Oriented Critical Review of Scientific Reliability and Future Research Directions

Aswin Karkadakattil [a++*]

[a] *Advanced Manufacturing, Artificial Intelligence and Space Systems, Kerala, India.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: https://doi.org/10.9734/iaarj/2026/v8i1123

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/162019

Review Article

Received: 17/05/2026
Accepted: 06/07/2026
Published: 11/07/2026

Abstract

The rapid expansion of large-scale astronomical surveys, multi-messenger observations, and high-resolution numerical simulations has transformed astronomy and astrophysics into increasingly data-intensive sciences, driving the widespread adoption of artificial intelligence (AI) and machine learning (ML) for scientific analysis. Machine learning has achieved notable success in applications including stellar parameter estimation, galaxy morphology classification, exoplanet detection, gravitational-wave analysis, and cosmological inference. However, conventional data-driven approaches often rely on statistical correlations, limiting their physical interpretability, robustness under distribution shifts, and ability to generalise to sparsely sampled or previously unseen astrophysical conditions. These challenges have stimulated growing interest in physics-informed machine learning (PIML), which integrates physical laws, governing equations,

[++] *Independent Researcher;*
**Corresponding author: E-mail: ashwinharik20000@gmail.com;*

Cite as: *Karkadakattil, A. (2026). Physics-Informed Machine Learning in Astronomy and Astrophysics: A Methodology-Oriented Critical Review of Scientific Reliability and Future Research Directions. International Astronomy and Astrophysics Research Journal, 8(1), 13–48. https://doi.org/10.9734/iaarj/2026/v8i1123*

conservation principles, and simulation-based knowledge into the learning process to improve scientific consistency and predictive reliability. This methodology-oriented critical review examines recent developments in PIML across major areas of astronomy and astrophysics, including stellar evolution, galaxy formation, cosmology, gravitational-wave astronomy, solar and space weather forecasting, and astrophysical fluid dynamics. Rather than providing an application-based survey, the review introduces a unified classification framework that organises existing methodologies according to their degree of physics integration, encompassing physics-guided learning, physics-constrained learning, physics-informed neural networks, neural operators, hybrid scientific machine learning, and emerging digital twin concepts. Using this framework, the literature is critically evaluated in terms of predictive capability, physical consistency, interpretability, computational scalability, uncertainty quantification, and methodological maturity. The review identifies several challenges that continue to limit broader adoption, including the scarcity of high-quality observational data, the computational cost of physics-constrained optimisation, the absence of standardised benchmarking practices, and limited validation under out-of-distribution astrophysical conditions. It also discusses emerging research directions, such as neural operators, multi-fidelity learning, uncertainty-aware AI, scientific foundation models, and digital twins, while emphasising that many of these technologies remain at an early stage of development for astronomical applications and require rigorous validation. Overall, the analysis indicates that increasing the integration of physical knowledge generally enhances scientific reliability, interpretability, and extrapolation capability, although these improvements are accompanied by greater computational complexity and methodological challenges. By providing a structured, methodology-oriented synthesis of the current literature, this review offers a critical perspective on the evolution of physics-informed AI and outlines key directions for developing scientifically reliable, interpretable, and physically grounded machine learning frameworks for next-generation astronomy and astrophysics.

Keywords: Physics-informed machine learning; scientific machine learning; astronomy; astrophysics; physics-informed neural networks; neural operators; explainable artificial intelligence; uncertainty quantification; digital twins; scientific foundation models.

1. Introduction

1.1 Astronomy as a Data-Intensive Science

Over the past two decades, astronomy and astrophysics have evolved into data-intensive sciences driven by next-generation observatories, large-scale sky surveys, high-fidelity numerical simulations, and advances in high-performance computing. Modern facilities, including the Legacy Survey of Space and Time (LSST), *Gaia*, *Euclid*, the James Webb Space Telescope (JWST), and the Square Kilometre Array (SKA), are producing unprecedented volumes of multi-wavelength imaging, spectroscopy, time-domain observations, and radio interferometric data. Together with cosmological, magnetohydrodynamic, and stellar evolution simulations, these resources have fundamentally changed how astronomical research is conducted (Ball & Brunner, 2010; Fluke & Jacobs, 2020; Sen et al., 2022; Ting, 2025). The emergence of multi-messenger astronomy has further increased data complexity by combining electromagnetic observations with gravitational waves, neutrinos, and cosmic rays to investigate extreme astrophysical phenomena. Consequently, astronomical datasets are characterised not only by their volume but also by their high dimensionality, heterogeneity, and temporal variability. Conventional analysis techniques based on manual feature engineering, empirical modelling, and statistical regression are increasingly unable to process these datasets efficiently or extract their full scientific value (Ball & Brunner, 2010; Baron, 2019; Sen et al., 2022). Artificial intelligence (AI), particularly machine learning (ML), has therefore become an essential component of modern astronomical research. ML algorithms now support a wide range of applications, including galaxy morphology classification, stellar parameter estimation, exoplanet detection, transient identification, gravitational-wave signal analysis, cosmological parameter inference, and space-weather forecasting (Meher & Panda, 2021; Li et al., 2025; Ting, 2025). These approaches have substantially accelerated data processing and enabled near-real-time analysis of observations that would otherwise require extensive computational effort. Despite these advances, most current ML models remain predominantly data-driven. Their optimisation is guided by statistical objectives rather than astrophysical principles, making them vulnerable to observational bias, reduced interpretability, and limited generalisation beyond the training distribution. Models may achieve excellent predictive accuracy while generating physically inconsistent results when applied to rare events or previously unseen astrophysical environments. These shortcomings have highlighted the need for learning frameworks that combine the predictive capability of machine learning with the physical principles governing astronomical systems. Recent guidance on machine learning project design in astronomy further emphasises that clear problem definition, model verification, characterisation, calibration and uncertainty-aware workflows are necessary for obtaining scientifically useful results (Buchner & Fotopoulou, 2024).

1.2 Evolution of Artificial Intelligence in Astronomy

The development of AI in astronomy has progressed through successive methodological stages. Early applications relied primarily on statistical inference, Bayesian methods, and empirical regression to estimate physical parameters and classify astronomical objects (Ball & Brunner, 2010). As observational datasets expanded, conventional machine learning algorithms including support vector machines, random forests, gradient boosting methods, and artificial neural networks became widely adopted for automated classification and regression tasks (Baron, 2019; Fluke & Jacobs, 2020). The introduction of deep learning further transformed astronomical data analysis by enabling automatic feature extraction from images, spectra, and time-series observations. Convolutional neural networks, recurrent neural networks, graph neural networks, and transformer architectures have achieved state-of-the-art performance in applications such as stellar spectroscopy, galaxy morphology classification, transient detection, and gravitational-wave analysis (Meher & Panda, 2021; Li et al., 2025; Ting, 2025). Deep learning has also enabled rapid estimation of stellar properties from asteroseismic observations and large stellar surveys (Bellinger et al., 2016; Mombarg et al., 2021). However, high predictive performance does not necessarily imply scientific reliability. Conventional deep learning models generally minimise statistical loss functions without enforcing conservation laws, governing differential equations, or other astrophysical constraints. Consequently, they often exhibit poor extrapolation under sparse observations, limited physical interpretability, and reduced robustness outside the training domain. These limitations are particularly significant in astronomy, where observations are frequently incomplete, noisy, or impossible to reproduce experimentally. Physics-informed machine learning (PIML) has emerged as a promising solution to these challenges by embedding physical knowledge directly into the learning process through governing equations, conservation laws, simulation-derived constraints, and domain expertise (Hao et al., 2022; Meng et al., 2025; Ghalambaz et al., 2024). Depending on the application, physical information may be incorporated through feature engineering, customised loss functions, hybrid simulation–data frameworks, or physics-informed neural networks (PINNs). Recent studies have successfully applied these approaches to solar magnetic field reconstruction (Jarolim et al., 2023), astrophysical shock modelling (Moschou et al., 2023), neutron-star magnetospheres (Urbán et al., 2023), dark-matter hydrodynamic simulations (Dai et al., 2024), neutrino flavour evolution (Abbar et al., 2024), stellar modelling (Baty, 2023), and magnetic classification of solar active regions (Li et al., 2026). These developments indicate a broader transition from prediction-oriented AI towards scientifically informed machine learning capable of integrating observations, simulations, and physical theory within unified computational frameworks. Recent reviews of unsupervised learning and variable-star research further show that automated feature discovery, representation learning and multimodal analysis are becoming important components of astronomical workflows, particularly for large survey and time-domain datasets (Fotopoulou, 2024; Audenaert, 2025). Methodological developments in neighbouring computational sciences also indicate that physics-guided learning, PINN-based inverse modelling and AI-augmented numerical schemes are increasingly discussed in relation to limited-label data, domain shift, uncertainty quantification and computationally demanding differential-equation problems, which are closely aligned with the reliability concerns considered in this review (Zhai, 2026; Okwuwe & Oduselu-Hassan, 2024).

1.3 Why Another Review?

Recent reviews have documented the rapid adoption of artificial intelligence (AI) and machine learning (ML) across astronomy. Early studies established the foundations of machine learning for astronomical data analysis (Ball & Brunner, 2010), while subsequent reviews examined its practical implementation and growing maturity within the field (Baron, 2019; Fluke & Jacobs, 2020). More recent surveys have highlighted the expansion of AI-driven methodologies across diverse astronomical applications and the increasing availability of large observational datasets (Sen et al., 2022; Ting, 2025). In parallel, broader reviews have examined the theoretical foundations of physics-informed machine learning (PIML) in computational science and engineering (Hao et al., 2022; Meng et al., 2025; Wayo, 2026). Although these studies provide valuable perspectives, they generally focus either on astronomical applications or on generic PIML methodologies, leaving limited discussion of how physics-informed learning has evolved specifically within astronomy and astrophysics. More importantly, existing reviews rarely examine when and why different levels of physics integration are beneficial. Questions such as when physics-guided learning is sufficient, when fully physics-informed models are required, how physical constraints influence robustness and generalisation, and what trade-offs exist between computational cost and scientific benefit remain insufficiently addressed. In addition, emerging developments including neural operators, scientific foundation models, uncertainty-aware learning, digital twins, and autonomous scientific AI have not yet been synthesised within a unified astronomy-focused framework. This review addresses these gaps through a methodology-oriented rather than application-oriented perspective. Instead of cataloguing individual studies, the literature is systematically classified according to the degree of physics integration and critically evaluated across major astronomical domains. The principal contribution is a unified framework that assesses AI methodologies based on their level of physics integration and scientific maturity rather than predictive performance alone. Building on this framework, the review compares the strengths and limitations of current approaches, identifies unresolved methodological challenges, and proposes a future research roadmap spanning neural operators, multi-fidelity learning, scientific foundation models, digital twins, uncertainty-aware AI, and autonomous scientific discovery. The conceptual evolution from conventional data-driven machine learning to trustworthy scientific AI, together with the increasing integration of physical knowledge across successive methodological stages, is illustrated in Fig. 1.

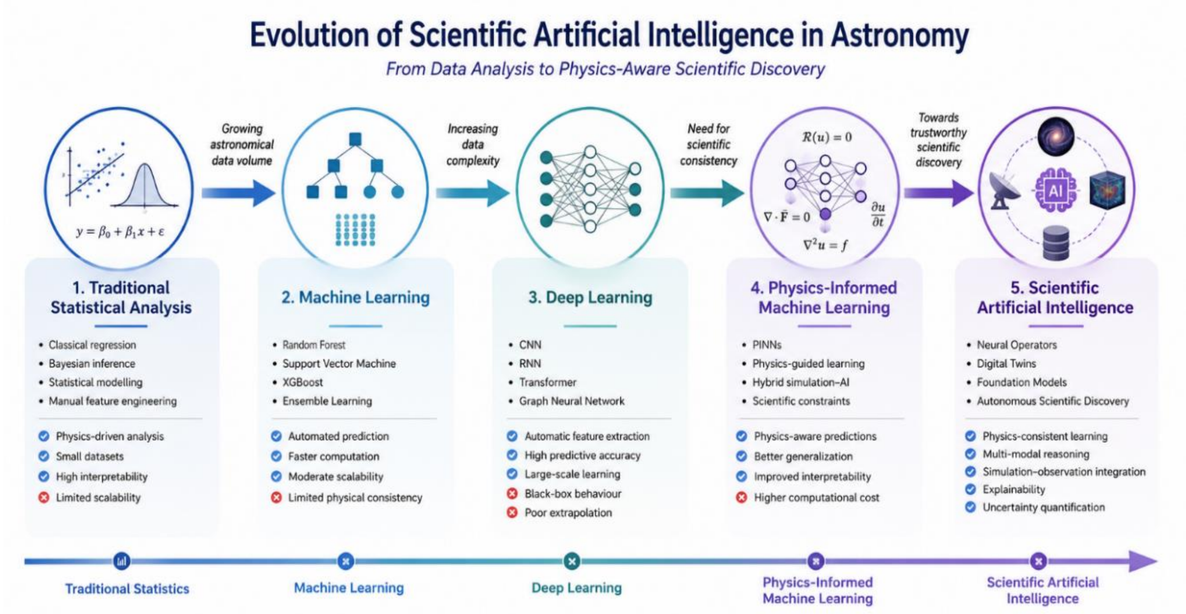

Fig. 1. Evolution of artificial intelligence methodologies in astronomy and astrophysics, illustrating the transition from traditional statistical analysis to machine learning, deep learning, physics-informed machine learning (PIML), and scientific artificial intelligence (Scientific AI)

The framework highlights the progressive integration of physical knowledge into data-driven models as astronomical datasets increase in scale and complexity. While conventional machine learning and deep learning primarily optimize statistical performance, PIML incorporates governing equations, physical constraints, and simulation knowledge to improve physical consistency, robustness, and generalization. The emerging Scientific AI paradigm further integrates observations, numerical simulations, neural operators, digital twins, foundation models, uncertainty quantification, and explainable AI to enable trustworthy and physically consistent scientific discovery. Figure developed by the author based on the synthesis of the literature (Ball and Brunner, 2010; Fluke and Jacobs, 2020; Hao et al., 2022; Meng et al., 2025; Ting, 2025).

1.4 Objective of the Review

The objective of this review is to critically synthesise physics-informed machine learning methodologies in astronomy and astrophysics through a methodology-oriented framework, compare their levels of physics integration and scientific reliability, identify current limitations, and outline research directions for trustworthy and physically grounded scientific artificial intelligence.

2. Review Methodology

2.1 Review Design and Literature Search Strategy

This review was designed as a methodology-oriented critical review supported by a structured and systematic literature search. Unlike a quantitative systematic review or meta-analysis, the primary objective of this study was not to exhaustively catalogue every published article but to critically evaluate the evolution of physics-informed machine learning (PIML) methodologies in astronomy and astrophysics, identify methodological trends, assess scientific reliability, and highlight emerging research directions. A structured review protocol was adopted to improve transparency, minimise selection bias, and ensure balanced coverage of both established and emerging developments. The literature search was conducted between April and May 2026 using multiple internationally recognised scientific databases, including NASA Astrophysics Data System (ADS), Scopus, Web of Science Core Collection, IEEE Xplore, SpringerLink, ScienceDirect, Google Scholar, and the arXiv preprint repository. These databases collectively provide comprehensive coverage of astronomy, astrophysics, computational physics, artificial intelligence, and scientific machine learning research. The primary search period covered publications from 2018 to 2026, corresponding to the rapid emergence of physics-informed machine learning and scientific artificial intelligence. Earlier landmark publications were also included where necessary to provide historical context regarding the evolution of machine learning in astronomy and astrophysics. Search queries combined astronomy-specific and machine-learning-specific keywords using Boolean operators (AND/OR). Representative search terms included "physics-informed machine learning," "physics-informed neural networks," "scientific machine learning," "machine learning in astronomy," "deep learning in astrophysics," "graph neural networks," "stellar evolution," "galaxy formation," "cosmological simulations," "gravitational waves," "space weather," "neural operators," "digital twins," "scientific foundation models," "surrogate modelling," and "explainable artificial intelligence." Forward citation tracking, backward reference screening, and cross-referencing of influential review articles were additionally employed to identify relevant publications that may not have been retrieved through database searches alone. Following the initial literature search, retrieved publications were screened for scientific relevance, methodological quality, originality, and direct applicability to astronomy and astrophysics. Rather than aiming for exhaustive coverage, emphasis was placed on selecting representative studies that collectively capture the major methodological developments within the field. The final review corpus consisted of 33 carefully selected publications, including landmark studies, influential review papers, and recent methodological advances that form the basis of the critical analyses presented throughout this review.

2.2 Study Selection and Quality Assessment

A predefined set of inclusion and exclusion criteria was used to ensure consistency during literature selection and to maintain the scientific quality of the reviewed studies.

Studies were included if they:

- addressed astronomy, astrophysics, cosmology, or closely related space science applications;
- proposed, applied, or critically reviewed machine learning, deep learning, scientific machine learning, or physics-informed learning methodologies;
- presented sufficient methodological detail to enable critical evaluation;
- provided original methodological developments, comprehensive reviews, or significant application studies; and
- were published in peer-reviewed journals or, where appropriate, well-recognised preprint repositories representing emerging research directions.

Studies were excluded if they:

- focused exclusively on engineering or other non-astronomical applications without methodological relevance to astronomy;
- consisted solely of editorials, opinion articles, conference abstracts, or short communications lacking sufficient technical detail;
- duplicated previously published work; or
- contained insufficient methodological information for meaningful critical assessment.

Rather than selecting studies solely on the basis of publication count or citation metrics, the literature was evaluated according to multiple quality indicators, including methodological originality, degree of physics integration, validation strategy, computational framework, scientific reproducibility, practical applicability, and potential impact on future astronomical research. This selection strategy ensured balanced coverage of both established methodologies and emerging research directions while maintaining a strong emphasis on scientific quality.

2.3 Methodology-Oriented Classification Framework

A distinguishing feature of the present review is its methodology-oriented organization, which differs from the conventional application-based structure adopted by most previous review articles. Existing surveys typically classify studies according to scientific domains such as stellar evolution, exoplanet detection, galaxy formation, cosmology, or space weather. While such classifications provide useful summaries of application areas, they offer limited insight into how different machine learning methodologies incorporate physical knowledge or how varying levels of physics integration influence scientific reliability, interpretability, computational efficiency, and generalisation capability. To address this limitation, the reviewed literature was systematically classified according to the degree of physics integration within the learning framework. Five progressively increasing levels of physics integration were identified.

Level 1 – Pure Data-Driven Machine Learning: Conventional machine learning algorithms, including random forests, support vector machines, gradient boosting methods, artificial neural networks, and deep neural networks, that learn statistical relationships directly from observational or simulated datasets without explicitly incorporating physical knowledge.

Level 2 – Physics-Guided Learning: Approaches that indirectly incorporate physical knowledge through feature engineering, physically meaningful descriptors, theoretical priors, domain-informed preprocessing, or expert-designed input variables while retaining conventional statistical optimisation.

Level 3 – Physics-Constrained Learning: Methods that directly integrate physical principles into model optimisation through customized loss functions, conservation laws, regularization strategies, boundary conditions, or hybrid simulation-data constraints, thereby improving physical consistency without completely reformulating the learning algorithm.

Level 4 – Scientific Machine Learning: Integrated computational frameworks that combine governing equations, numerical simulations, observational data, and machine learning within unified learning architectures. Representative approaches include Physics-Informed Neural Networks (PINNs), neural operators, physics-informed surrogate models, and other scientific machine learning techniques capable of simultaneously exploiting observational and physical information.

Level 5 – Emerging Scientific Digital Twins: Adaptive computational frameworks that combine continuous observational data assimilation, numerical simulations, explainable artificial intelligence, uncertainty quantification, and adaptive machine learning to construct dynamic digital representations of astrophysical systems. Although these approaches show considerable promise, practical astronomical implementations remain relatively limited, and additional methodological validation is required before widespread scientific adoption. This hierarchical framework enables direct comparison of machine learning methodologies according to their degree of physics integration, scientific maturity, computational complexity, interpretability, robustness, and potential contribution to reliable scientific discovery across diverse astronomical applications.

2.4 Critical Evaluation Framework

Rather than presenting the literature chronologically, the selected studies were critically evaluated using a multidimensional assessment framework specifically developed for this review. The objective was to compare methodologies based on their scientific strengths, limitations, and practical applicability rather than predictive performance alone. This methodology-oriented framework builds upon principles commonly adopted in recent reviews of physics-informed machine learning and scientific machine learning while extending them to the specific requirements of astronomy and astrophysics (Hao et al., 2022; Meng et al., 2025). The overall review methodology, including the structured literature search strategy, representative study selection process, methodology-oriented classification, and critical evaluation framework adopted throughout this review, is summarised in Fig. 2.

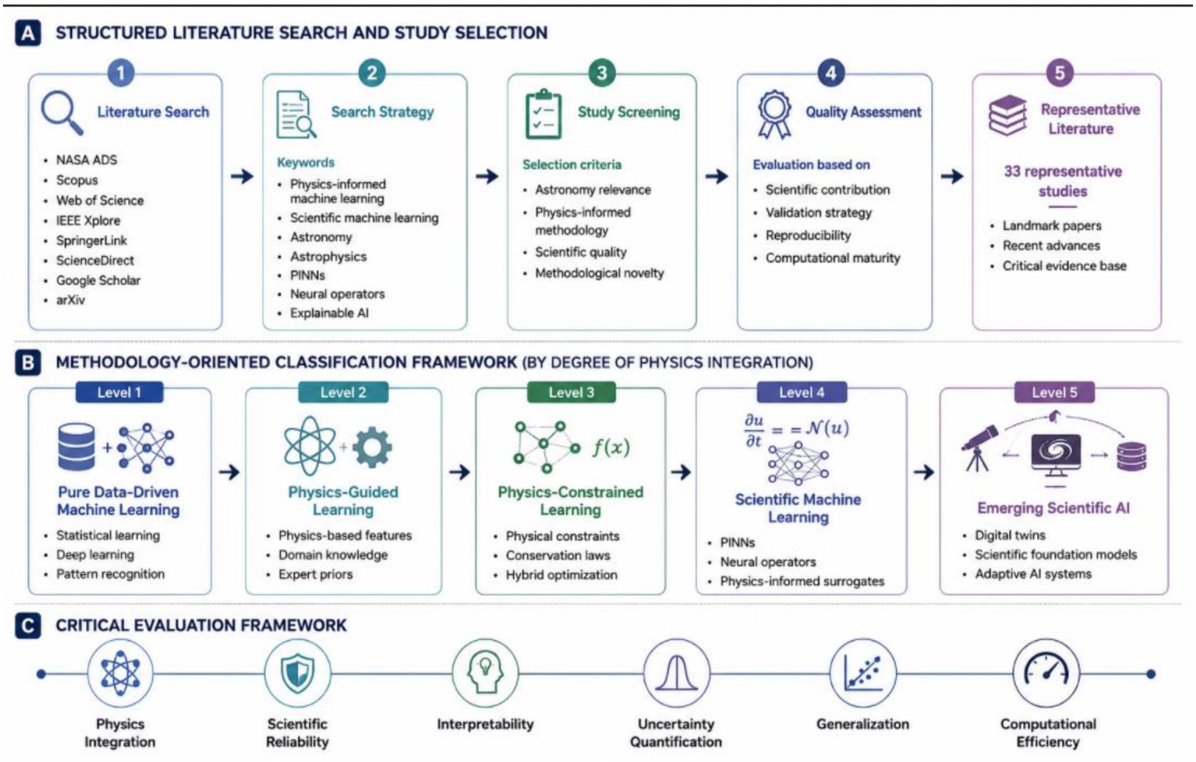

Fig. 2. Methodology-oriented framework adopted in this review, illustrating the structured literature search strategy, representative study selection, hierarchical classification of physics integration in machine learning methodologies, and the multidimensional evaluation framework used to critically assess scientific reliability, interpretability, computational efficiency, uncertainty quantification, and applicability across astronomy and astrophysics

Five complementary evaluation criteria were employed:

(i) Degree of physics integration: the extent to which physical laws, governing equations, conservation principles, or domain knowledge were incorporated into the learning framework.

(ii) Scientific reliability: the ability of the methodology to produce physically consistent predictions, maintain robustness under sparse or noisy observations, and generalise beyond the training domain.

(iii) Computational efficiency and scalability: the computational requirements associated with model training and inference, scalability to large astronomical datasets, and suitability for high-performance scientific computing.

(iv) Interpretability and uncertainty quantification: the transparency of model predictions, incorporation of explainable artificial intelligence techniques, treatment of predictive uncertainty, and overall scientific trustworthiness.

(v) Astronomical applicability and methodological maturity: the breadth of successful applications across astronomy and astrophysics, validation using observational or simulation data, reproducibility, and readiness for practical scientific deployment.

Applying a common evaluation framework enabled consistent comparison across diverse machine learning paradigms while shifting the emphasis from predictive accuracy alone toward broader measures of scientific quality, physical consistency, interpretability, and practical usefulness. Similar multidimensional perspectives have been advocated in recent reviews of scientific machine learning; however, the present framework extends these concepts by explicitly incorporating astronomy-specific considerations, including methodological maturity and applicability across diverse astrophysical domains (Hao et al., 2022; Meng et al., 2025; Ting, 2025). The resulting framework forms the analytical basis for the comparative discussions, methodological synthesis, and future research roadmap presented throughout the remainder of this review.

To provide a structured comparison of representative studies, Table 1 summarises the literature according to the astronomy domain, level of physics integration, methodological approach, validation strategy, key scientific contribution, and remaining limitations.

Table 1. Representative Literature on Physics Integration in Machine Learning for Astronomy and Astrophysics

Reference	Astronomy Domain	Physics Integration Level	Representative Method	Validation Strategy	Key Scientific Contribution	Remaining Limitation
Ball and Brunner (2010)	General astronomy	Level 1 – Data-driven ML	Classical machine learning	Astronomical survey datasets	Established machine learning as a practical tool for large-scale astronomical data mining	No incorporation of physical constraints or scientific interpretability
Fluke and Jacobs (2020)	General astronomy	Level 1 – Data-driven ML	Systematic review	Literature synthesis	Evaluated the maturity and adoption of AI across astronomy	Limited discussion of physics-informed methodologies
Jarolim et al. (2023)	Solar physics	Level 4 – Scientific ML	Physics-informed neural networks (PINNs)	Solar observations	Reconstructed coronal magnetic fields while enforcing physical consistency	High computational cost and problem-specific training
Moschou et al. (2023)	Astrophysical fluid dynamics	Level 4 – Scientific ML	PINNs	Numerical simulations	Modelled astrophysical shock propagation using governing physical equations	Scalability to complex multi-physics systems remains challenging
Urbán et al. (2023)	Neutron-star astrophysics	Level 4 – Scientific ML	PINNs	Numerical solutions	Modelled force-free neutron-star magnetospheres with improved physical consistency	Computational efficiency decreases for large-scale simulations
Dai et al. (2024)	Cosmology	Level 4 – Scientific ML	PINNs	Hydrodynamic simulations	Reconstructed baryonic matter distributions from dark-matter simulations	Generalisation across different cosmological simulations requires further validation
Li et al. (2026)	Solar physics	Level 3 – Physics-constrained learning	Physics-informed deep learning	Solar observational datasets	Improved magnetic classification of solar active regions using embedded physical knowledge	Limited validation under unseen solar activity conditions

Reference	Astronomy Domain	Physics Integration Level	Representative Method	Validation Strategy	Key Scientific Contribution	Remaining Limitation
Sun et al. (2025)	Digital twins	Level 5 – Digital twins	Generative AI with Bayesian updating	Review and conceptual demonstrations	Demonstrated the potential of Bayesian updating for adaptive digital twins	Applications to astronomy remain largely conceptual
Zhou et al. (2026)	Scientific AI	Level 5 – Autonomous Scientific AI	Digital Twin AI	Conceptual framework	Proposed AI-driven digital twins integrating large language models with scientific workflows	Practical astronomical implementations have yet to be demonstrated

Critical observation: The literature reveals a clear methodological evolution from purely statistical machine learning towards increasingly physics-aware scientific AI. While Level 1 approaches provide excellent computational efficiency, they frequently lack physical consistency and reliable extrapolation. Level 3 and Level 4 methodologies improve scientific robustness by embedding physical knowledge directly into model development, although this is achieved at the expense of greater computational complexity. Level 5 approaches, including digital twins and autonomous scientific AI, represent the emerging frontier of the field but remain at an early stage of practical deployment. This progression provides the conceptual basis for the methodology-oriented framework developed throughout this review.

3. From Black-Box Artificial Intelligence to Scientific Machine Learning

Artificial intelligence has become an integral part of modern astronomy, enabling efficient analysis of the massive observational and simulation datasets generated by contemporary astronomical facilities. Early machine learning methods demonstrated remarkable success in tasks such as source classification, stellar parameter estimation, transient detection, and galaxy morphology analysis by learning complex patterns directly from data (Ball & Brunner, 2010; Fluke & Jacobs, 2020; Ting, 2025). However, these models are largely data-driven and often provide limited insight into the physical mechanisms underlying their predictions. As AI has evolved from a data-analysis tool to a component of scientific investigation, increasing emphasis has been placed on integrating physical knowledge into the learning process. This has led to the emergence of scientific machine learning, where observational data are combined with governing equations, numerical simulations, and domain expertise to improve physical consistency, robustness, and interpretability (Hao et al., 2022; Meng et al., 2025; Ghalambaz et al., 2024). Rather than treating machine learning and physics as competing approaches, current research increasingly combines their complementary strengths to develop models that are both computationally efficient and scientifically reliable. The conceptual evolution of these methodologies, from conventional data-driven learning to increasingly physics-aware scientific AI frameworks, is illustrated in Fig. 3. The following sections critically examine this methodological progression, spanning conventional machine learning, physics-guided learning, physics-constrained models, Physics-Informed Neural Networks (PINNs), neural operators, and emerging scientific AI frameworks. Each approach is evaluated in terms of its level of physics integration, scientific strengths, computational limitations, and suitability for different astronomical applications.

The conceptual framework illustrates the progression from purely data-driven machine learning toward increasingly physics-aware paradigms, culminating in scientific machine learning and digital twin-based AI. As the degree of physics integration increases, models generally become more physically consistent and scientifically reliable, although they also tend to require greater computational resources and methodological complexity. This evolution reflects the broader transition from prediction-oriented AI to trustworthy scientific discovery in astronomy and astrophysics.

3.1 Pure Machine Learning

Pure machine learning (ML) represents the foundation of artificial intelligence in modern astronomy. These approaches learn statistical relationships directly from observational or simulated datasets without explicitly incorporating astrophysical knowledge into the learning process. Over the past decade, algorithms such as support vector machines, random forests, gradient boosting, artificial neural networks, and deep neural networks have become indispensable tools for analysing the enormous volumes of data produced by contemporary astronomical surveys. Their success stems from their ability to model highly nonlinear relationships that are often difficult or impossible to represent using conventional analytical methods (Ball & Brunner, 2010; Baron, 2019; Fluke & Jacobs, 2020). The widespread adoption of pure ML has been driven by the unprecedented growth of astronomical data. Modern observatories such as Gaia, JWST, Euclid, and the Legacy Survey of Space and Time (LSST) continuously generate petabyte-scale datasets containing multi-wavelength images, spectra, and time-series observations. Processing these datasets using traditional statistical techniques is computationally prohibitive, whereas machine learning enables rapid automated analysis with minimal human intervention. Consequently, ML has become central to numerous astronomical applications, including galaxy morphology classification, stellar parameter estimation, exoplanet detection, transient identification, gravitational-wave signal classification, cosmological parameter inference, and anomaly detection (Meher & Panda, 2021; Sen et al., 2022; Ting, 2025). Several landmark studies illustrate the effectiveness of purely data-driven learning. Bellinger et al. (2016) demonstrated that neural networks could estimate the fundamental parameters of main-sequence stars within seconds, replacing computationally intensive stellar modelling. Similarly, deep learning has significantly improved the interpretation of asteroseismic observations by enabling rapid estimation of stellar masses, ages, and internal mixing parameters from oscillation spectra (Mombarg et al., 2021). More recently, comprehensive reviews by Li et al. (2025) and Ting (2025) have shown that deep learning now underpins many state-of-the-art astronomical pipelines, particularly for large-scale imaging and spectroscopic surveys. Another important advantage of conventional ML is its computational efficiency after training. Modern deep neural networks can analyse millions of astronomical images or spectra within minutes, making them particularly suitable for real-time survey pipelines and transient-event detection. Automatic feature learning has also reduced dependence on manually engineered descriptors, allowing models to identify subtle patterns that may not be evident through traditional statistical analysis. These capabilities have substantially accelerated scientific workflows and enabled discoveries that would have been impractical using conventional computational approaches (Fluke & Jacobs, 2020; Meher & Panda, 2021). Despite these achievements, purely data-driven models exhibit several fundamental limitations that restrict their use as reliable scientific tools. The optimisation process is designed to minimise statistical prediction error rather than enforce astrophysical principles. Consequently, a model may achieve excellent predictive accuracy while simultaneously violating conservation laws, governing differential equations, or well-established physical relationships. High statistical performance should therefore not be interpreted as evidence of scientific correctness. This distinction is particularly important in astronomy, where observations are frequently sparse, noisy, and limited to a small portion of the underlying physical parameter space (Hao et al., 2022; Meng et al., 2025). Generalisation beyond the training distribution remains another significant challenge. Machine learning algorithms generally perform well when new observations resemble the training data but often degrade under distribution shifts. Astronomical research routinely investigates rare stellar populations, extreme astrophysical environments, and transient events that are poorly represented in existing datasets. Models trained on nearby stellar populations, for example, may produce unreliable predictions when applied to stars with different metallicities, evolutionary stages, or magnetic properties. Similar challenges arise in cosmology, where surrogate models calibrated using one numerical simulation often perform poorly when transferred to alternative cosmological scenarios or observational surveys (Villaescusa-Navarro et al., 2021; Kugel et al., 2023). Interpretability presents an additional obstacle. Most deep neural networks operate as black-box models, making it difficult to determine whether their predictions arise from genuine astrophysical relationships or from spurious statistical correlations. Explainable AI techniques such as SHAP and LIME improve transparency by identifying influential input variables, but they do not establish physical causality. A highly ranked feature may simply reflect biases within the training dataset rather than a meaningful astrophysical mechanism (Mohammadi et al., 2022; Kumaran et al., 2025). These limitations become increasingly important as machine learning transitions from a data-processing tool to a framework for scientific inference. While conventional ML has transformed astronomical data analysis through its speed, flexibility, and predictive capability, its limited physical interpretability, poor extrapolation, and inability to guarantee physically consistent predictions restrict its application in problems where scientific reliability is essential. Consequently, the next generation of astronomical AI increasingly seeks to integrate domain knowledge directly into the learning process, giving rise to physics-guided, physics-constrained, and physics-informed machine learning frameworks**.** Pure machine learning remains the most computationally mature and widely deployed approach in astronomy, particularly for large observational datasets with abundant labelled examples. However, its dependence on statistical correlations rather than physical reasoning fundamentally limits scientific interpretability and extrapolation. Rather than replacing conventional ML, subsequent generations of scientific AI seek to retain its computational efficiency while embedding physical knowledge to improve robustness, interpretability, and trustworthiness. This progression forms the basis for the evolution from black-box prediction towards scientifically informed machine learning discussed in the following sections.

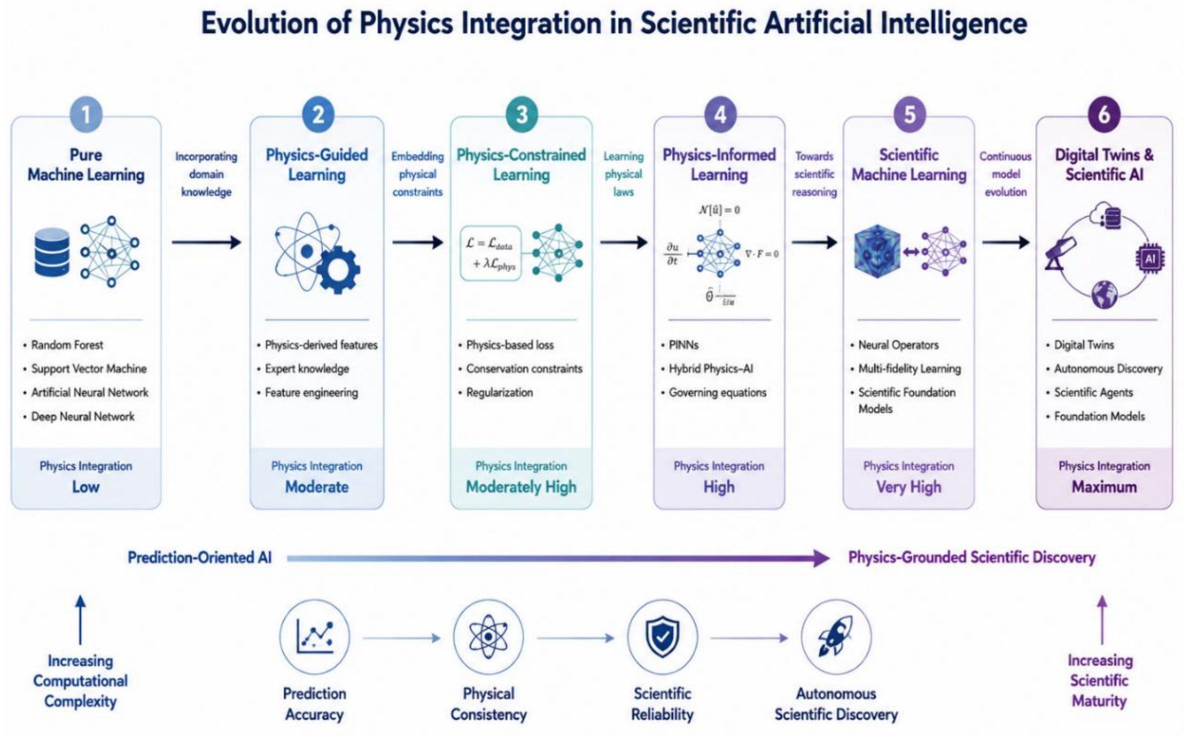

Fig. 3. Evolution of physics integration in scientific artificial intelligence

3.2 Physics-Guided Learning

Physics-guided learning represents an important transition between purely data-driven machine learning and fully physics-informed artificial intelligence. Rather than embedding governing equations directly into the optimisation process, these methods incorporate existing physical knowledge through feature engineering, theoretical priors, data preprocessing, or physically meaningful constraints on the model inputs and outputs. The underlying learning algorithm remains data-driven, but the information supplied to the model is enriched using well-established scientific understanding. This strategy attempts to retain the computational efficiency and flexibility of conventional machine learning while improving the physical relevance of its predictions (Pilania et al., 2018; Hao et al., 2022; Meng et al., 2025). The motivation for physics-guided learning is straightforward. Astronomical observations are often sparse, noisy, and affected by instrumental limitations, making it difficult for purely statistical models to distinguish physically meaningful relationships from spurious correlations. Incorporating domain knowledge into the learning process reduces the search space and directs the model towards physically plausible solutions. Instead of learning entirely from raw observational data, the algorithm is supplied with descriptors derived from established physical theory, numerical simulations, or empirical astrophysical relationships. In astronomy, physics-guided learning has been applied across a wide range of problems. Derived stellar quantities from evolutionary models, asteroseismic parameters, cosmological descriptors, radiative-transfer variables, and physically meaningful spectral indices are frequently incorporated as model inputs to improve predictive performance. Studies on stellar parameter estimation have demonstrated that physically motivated descriptors reduce prediction uncertainty while improving convergence compared with purely empirical feature sets (Bellinger et al., 2016; Mombarg et al., 2021). Similar strategies have been adopted in cosmological simulations, where theoretical quantities describing dark matter distribution or large-scale structure provide stronger physical guidance than raw simulation outputs alone (Villaescusa-Navarro et al., 2021). One of the principal advantages of physics-guided learning is its relatively low computational overhead. Because governing equations are not explicitly solved during optimisation, these methods remain computationally efficient and can often be implemented using existing machine learning architectures. Introducing physically meaningful features also reduces the volume of training data required, improves optimisation stability, and frequently enhances model interpretability because the selected variables possess direct physical significance rather than abstract statistical representations. Similar benefits have been reported in materials science, climate modelling, and computational physics, where physics-guided feature engineering improves prediction accuracy without substantially increasing computational complexity (Pilania et al., 2018; Meray et al., 2024). Despite these advantages, physics-guided learning does not fundamentally solve the scientific limitations of conventional machine learning. Physical information influences the model indirectly through engineered features or prior knowledge, but it is not enforced during optimisation. Consequently, the algorithm may still generate predictions that violate governing equations or conservation principles if doing so reduces the statistical loss function. In other words, physics guides the learning process but does not constrain it. As a result, predictive accuracy may improve while physical consistency remains uncertain, particularly when models are applied outside the range of the training data (Hao et al., 2022; Meng et al., 2025). Another challenge is the dependence on expert knowledge during feature construction. Designing informative physics-based descriptors requires substantial domain expertise and may introduce subjective bias into the modelling process. Furthermore, manually engineered features are often specific to individual applications and may not transfer effectively across different astronomical problems. This limitation becomes increasingly important for complex multi-physics systems where the governing interactions are only partially understood or involve coupled physical processes spanning multiple spatial and temporal scales. Physics-guided learning therefore represents an important methodological improvement rather than a complete solution. It demonstrates that incorporating physical knowledge can improve robustness, reduce data requirements, and enhance scientific interpretability without sacrificing computational efficiency. However, because physical constraints remain external to the optimisation process, these methods cannot guarantee physically consistent predictions. This limitation has motivated the development of physics-constrained learning, where physical laws are embedded directly within the training objective to achieve a closer integration between data-driven learning and established scientific theory. Critical Perspective. Physics-guided learning provides an effective balance between predictive performance and computational efficiency, making it attractive for many practical astronomical applications. Nevertheless, its dependence on engineered physical features means that scientific reliability ultimately depends on the quality of the prior knowledge supplied to the model. While it represents a significant advance over purely empirical learning, it should be regarded as an intermediate step towards fully physics-informed frameworks that explicitly enforce governing physical principles during optimisation.

3.3 Physics-Constrained Learning

Physics-constrained learning represents a significant advance beyond physics-guided machine learning by incorporating physical principles directly into the model optimisation process. Unlike physics-guided approaches, where physical knowledge is introduced indirectly through engineered features or prior information, physics-constrained models embed scientific constraints within the objective function. During training, the algorithm simultaneously minimises prediction error and penalties associated with violating conservation laws, constitutive relationships, or other physically meaningful constraints. This approach enables the model to preserve much of the flexibility of conventional machine learning while improving physical consistency and scientific reliability (Hao et al., 2022; Meng et al., 2025). The central idea behind physics-constrained learning is that not every statistically optimal solution is scientifically acceptable. In astronomy and astrophysics, observations are often incomplete, noisy, or sparsely sampled, allowing purely data-driven algorithms to learn relationships that fit the available data but contradict well-established physical principles. By embedding constraints derived from astrophysical theory into the optimisation process, the search space is restricted to solutions that remain consistent with existing scientific knowledge. Consequently, the model is encouraged to learn physically meaningful representations instead of relying solely on statistical correlations. Physical constraints can be incorporated in several ways depending on the application. Conservation of mass, momentum, energy, or magnetic flux may be enforced through penalty terms added to the loss function. Other studies introduce astrophysical scaling relations, boundary conditions, equilibrium requirements, or theoretical consistency checks during optimisation. Hybrid simulation–data frameworks also belong to this category, where observational data guide the learning process while numerical simulations provide additional physical constraints that regularise model behaviour (Hao et al., 2022; Ghalambaz et al., 2024). Compared with conventional machine learning, physics-constrained learning generally produces models that are more robust under noisy observations and better suited for extrapolation beyond the training dataset. Because the optimisation process discourages physically unrealistic solutions, predictions are less sensitive to observational uncertainty and often remain stable when applied to conditions not extensively represented during training. These characteristics are particularly valuable in astronomy, where rare astrophysical events and extreme physical environments frequently lie outside the range of available observations. Another important advantage is improved scientific interpretability. Since the optimisation process explicitly considers physical consistency, the resulting predictions are easier to reconcile with existing astrophysical theory. This improves confidence in model outputs and facilitates their integration into scientific workflows, where understanding the underlying physical mechanisms is often as important as achieving high predictive accuracy. Despite these advantages, physics-constrained learning also presents important challenges. Designing appropriate constraint functions requires substantial domain expertise and depends strongly on the underlying physical problem. If the imposed constraints are incomplete, overly restrictive, or inconsistent with the available observations, model performance may deteriorate. Furthermore, balancing prediction accuracy against constraint satisfaction often requires careful tuning of weighting coefficients within the loss function. Excessive emphasis on physical constraints may reduce predictive flexibility, whereas insufficient weighting may fail to eliminate physically unrealistic solutions. This trade-off remains an active area of research across many scientific disciplines (Hao et al., 2022; Meng et al., 2025). Computational cost also increases relative to purely data-driven machine learning because additional physical constraints must be evaluated during each optimisation step. Although this increase is generally smaller than that associated with fully physics-informed neural networks, it becomes increasingly significant for coupled multi-physics problems involving numerous governing equations or high-dimensional parameter spaces. Consequently, the practical implementation of physics-constrained learning requires balancing scientific fidelity against computational efficiency. Physics-constrained learning should therefore be viewed as a bridge between conventional machine learning and fully physics-informed modelling. It demonstrates that integrating physical laws directly into optimisation substantially improves scientific reliability without requiring the complete reformulation of the learning algorithm. At the same time, it highlights the limitations of constraint-based approaches, particularly for complex systems governed by coupled partial differential equations. These challenges have motivated the development of Physics-Informed Neural Networks (PINNs), which integrate governing equations directly into the neural network architecture and optimisation framework.

Critical Perspective. Physics-constrained learning provides an effective compromise between computational efficiency and physical fidelity. Compared with physics-guided approaches, it offers stronger guarantees of scientific consistency and improved robustness under sparse or noisy observations. However, its performance depends critically on the selection and weighting of physical constraints, and it does not fully resolve the optimisation and scalability challenges encountered in more complex multi-physics systems. Consequently, physics-constrained learning represents an important transitional stage in the evolution of scientific machine learning rather than the final solution for physically consistent artificial intelligence.

3.4 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) have emerged as one of the most influential developments in scientific machine learning by embedding governing physical equations directly into the learning process. Unlike conventional neural networks, which optimise only against observational data, PINNs minimise a composite loss function that combines data mismatch with residuals of governing differential equations, boundary conditions, and initial conditions. This enables the model to satisfy both empirical observations and established physical laws, making PINNs particularly valuable for scientific problems where observations are sparse but the underlying physics is well understood (Hao et al., 2022; Ghalambaz et al., 2024; Meng et al., 2025). The adoption of PINNs has been driven by several challenges unique to astronomy and astrophysics. Many astrophysical processes cannot be reproduced experimentally, observational datasets are often incomplete or affected by measurement uncertainty, and high-fidelity numerical simulations are computationally expensive. By integrating physical constraints with observational data, PINNs reduce dependence on large labelled datasets and enable physically plausible solutions even under limited observational conditions. Consequently, they have become particularly attractive for inverse problems, parameter estimation, and data assimilation, where observational information alone is insufficient to uniquely determine the underlying physical state (Hao et al., 2022). Recent studies demonstrate the versatility of this approach across astronomy. Jarolim et al. (2023) reconstructed three-dimensional solar coronal magnetic fields while enforcing magnetohydrodynamic constraints, producing solutions that were more physically consistent than conventional extrapolation methods. PINNs have also been applied to astrophysical shock modelling (Moschou et al., 2023), Lane–Emden equations describing stellar interiors (Baty, 2023), neutron-star magnetospheres (Urbán et al., 2023), hydrodynamic reconstruction from dark matter distributions (Dai et al., 2024), and fast neutrino flavour conversion in core-collapse supernovae (Abbar et al., 2024). These studies illustrate that PINNs are evolving from a methodological concept into a practical computational tool for solving a wide range of astrophysical inverse and forward problems. A major advantage of PINNs is their ability to incorporate prior physical knowledge directly into model optimisation. Governing equations effectively act as an additional source of supervision, reducing the space of admissible solutions and improving physical plausibility when observational data are sparse or noisy. This characteristic is particularly important in astronomy, where telescope time is limited, rare events are difficult to observe, and many physical processes cannot be measured directly. Compared with purely data-driven models, PINNs often produce solutions that are more consistent with established physical theory and less sensitive to observational incompleteness. However, these advantages come with significant computational trade-offs. Training a PINN is substantially more expensive than training a conventional neural network because the optimisation process repeatedly evaluates governing equations and their derivatives throughout training. As the complexity of the governing physics increases, optimisation becomes slower, memory requirements increase, and convergence becomes more difficult (Hao et al., 2022; Ghalambaz et al., 2024). These challenges are particularly evident for nonlinear multi-physics systems, turbulent flows, magnetohydrodynamics, and large three-dimensional astrophysical simulations. Optimisation remains another important limitation. PINNs minimise several competing objectives simultaneously, including observational error, equation residuals, and boundary or initial conditions. Selecting appropriate weights for these loss components remains largely problem-dependent and often requires extensive hyperparameter tuning. Poorly balanced optimisation may converge to solutions that satisfy one objective while compromising others, reducing both accuracy and physical fidelity (Meng et al., 2025). Scalability also limits the routine application of PINNs to large astronomical problems. Although they perform well for many low- and moderate-dimensional systems, extending them to high-dimensional, time-dependent, or multi-scale simulations remains computationally demanding. Recent advances—including adaptive sampling, domain decomposition, curriculum learning, and hybrid PINN architectures—have improved training efficiency, but further methodological developments are required before PINNs can be routinely deployed for exascale astrophysical simulations (Ghalambaz et al., 2024). An important question is whether PINNs should replace conventional machine learning. Current evidence suggests that the answer is **no**. When large, high-quality labelled datasets are available and the objective is efficient prediction or classification, conventional deep learning often achieves comparable predictive performance with substantially lower computational cost. In contrast, PINNs are most beneficial when reliable governing equations exist, observations are sparse or noisy, or the scientific objective requires solutions that remain consistent with known physical laws. Consequently, the choice between conventional machine learning and PINNs should be driven by the characteristics of the scientific problem rather than by algorithmic preference.

Critical Perspective: PINNs represent a significant advance in the transition from empirical machine learning to physics-aware scientific AI. Their greatest contribution is not necessarily higher predictive accuracy but the ability to integrate physical knowledge directly into the learning process, improving the credibility of predictions under data-limited conditions. Nevertheless, high computational cost, optimisation instability, and limited scalability remain important barriers to widespread adoption. Future research is therefore likely to focus on hybrid frameworks that combine PINNs with neural operators, surrogate models, and foundation models to achieve a more effective balance between physical fidelity, computational efficiency, and scalability.

3.5 Neural Operators

Neural operators have emerged as a major advance in scientific machine learning by learning mappings between **function spaces** rather than individual input–output pairs. Unlike conventional neural networks, which approximate specific solutions, neural operators learn the underlying solution operator of a differential equation, enabling them to predict solutions for new initial conditions, boundary conditions, or physical parameters without retraining. This capability addresses one of the principal limitations of Physics-Informed Neural Networks (PINNs), namely their limited scalability for repeated large-scale simulations. Instead of solving a governing equation for every new physical configuration, neural operators learn the relationship between physical inputs and solutions directly from simulation data. Once trained, they can rapidly generate predictions for previously unseen scenarios, making them particularly attractive for computationally intensive applications. Architectures such as the Fourier Neural Operator (FNO) and Deep Operator Network (DeepONet) have demonstrated excellent performance in fluid dynamics, heat transfer, climate modelling, and physics-based surrogate modelling, often achieving orders-of-magnitude faster inference than conventional numerical solvers while maintaining high predictive accuracy (Lu & Wang, 2024). Although their application in astronomy is still emerging, their potential is considerable. Many astrophysical problems including cosmological structure formation, radiative transfer, magnetohydrodynamics, and plasma physics require repeated solutions of complex partial differential equations. Neural operators can substantially accelerate these simulations, making them particularly valuable for uncertainty quantification, Bayesian parameter estimation, surrogate modelling, and large-scale parameter exploration. A key advantage of neural operators is computational scalability. Unlike PINNs, which optimise each new problem individually, neural operators perform the computationally intensive learning during an offline training stage. Once trained, inference is extremely efficient, enabling rapid evaluation of thousands of new physical configurations. This makes them well suited to next-generation astronomical workflows where repeated simulations dominate the overall computational cost. However, these advantages are accompanied by important limitations. Neural operators rely heavily on the availability of large, diverse, and high-quality simulation datasets. Generating such datasets may itself require substantial computational resources using traditional numerical solvers, shifting rather than eliminating the computational burden. Furthermore, because neural operators learn from simulation data instead of explicitly enforcing governing equations during prediction, their performance depends strongly on the quality and physical coverage of the training dataset. Predictions may become unreliable when applied beyond the simulated parameter space or to previously unseen physical regimes. Another important challenge is interpretability. While neural operators can accurately approximate complex physical processes, the learned operator often provides limited insight into the underlying physical mechanisms. Consequently, understanding why a particular prediction is produced remains more difficult than verifying whether it satisfies known physical laws. Integrating operator learning with uncertainty quantification, explainable AI, and physics-based constraints therefore represents an important direction for future research. Current evidence suggests that neural operators should not be viewed as replacements for PINNs but as complementary methodologies. PINNs are generally preferable when governing equations are well established and observational data are sparse, whereas neural operators are better suited to applications requiring rapid inference from large simulation datasets. The choice between these approaches should therefore be guided by the scientific objective, data availability, and computational requirements rather than by algorithmic preference alone.

Critical Perspective: Neural operators represent an important step towards scalable scientific AI by transferring much of the computational cost from repeated numerical optimisation to reusable operator learning. Their ability to accelerate large-scale simulations makes them particularly promising for computational astronomy. Nevertheless, their dependence on extensive simulation datasets and limited physical interpretability remain significant challenges. Future progress is likely to come from hybrid frameworks that combine neural operators with physics-informed learning, uncertainty quantification, and explainable AI, thereby achieving a more effective balance between computational efficiency, physical fidelity, and scientific reliability.

3.6 Hybrid Scientific Artificial Intelligence

The convergence of machine learning, numerical simulations, and physical modelling has given rise to a new generation of computational frameworks collectively referred to as Hybrid Scientific Artificial Intelligence (Scientific AI). Rather than relying exclusively on either observational data or theoretical models, these frameworks integrate multiple sources of scientific information to improve predictive accuracy, physical consistency, and computational efficiency. This represents a fundamental shift in the role of artificial intelligence within astronomy. Instead of functioning solely as a predictive tool, AI is increasingly becoming an integral component of scientific reasoning, capable of combining observations, simulations, and established physical knowledge within a unified computational framework. Hybrid Scientific AI recognises that neither data-driven learning nor physics-based modelling alone is sufficient to address many contemporary astrophysical challenges. Astronomical observations are frequently incomplete because of instrumental limitations, finite observing time, and the inability to reproduce most astrophysical phenomena experimentally. Numerical simulations provide physically consistent representations of these systems but are computationally demanding and often impractical for large parameter studies. Hybrid frameworks exploit the complementary strengths of both approaches. Observational data constrain model behaviour, numerical simulations provide physically realistic training examples, machine learning accelerates computation through surrogate modelling, and uncertainty quantification evaluates the reliability of model predictions. This integrated philosophy has already begun to influence several areas of computational astrophysics. Cosmological simulation projects such as CAMELS and FLAMINGO combine large-scale numerical simulations with machine learning to accelerate cosmological inference while preserving physical realism (Villaescusa-Navarro et al., 2021; Kugel et al., 2023).

Similarly, physics-informed surrogate models have demonstrated considerable potential for replacing computationally expensive numerical solvers in engineering and geoscience applications, suggesting comparable opportunities for radiative transfer, magnetohydrodynamics, stellar evolution, and galaxy formation studies (Meray et al., 2024; Lu & Wang, 2024). These developments indicate that future astronomical workflows are likely to depend increasingly on hybrid computational frameworks capable of combining simulation fidelity with machine learning efficiency. A key component of Hybrid Scientific AI is uncertainty quantification. Scientific predictions must provide not only accurate estimates but also reliable confidence measures that reflect observational uncertainty, model assumptions, and numerical approximations. Bayesian learning, probabilistic machine learning, ensemble methods, and physics-aware uncertainty estimation therefore play an increasingly important role in supporting robust scientific inference. Without reliable uncertainty estimates, even highly accurate predictions may be difficult to interpret within the context of hypothesis testing or observational planning. Another emerging direction is the integration of Explainable Artificial Intelligence (XAI) into scientific workflows. As machine learning models become increasingly complex, understanding why a prediction is produced becomes essential for scientific credibility. Recent studies have demonstrated that explainability techniques can reveal physically meaningful relationships in astronomical data, identify potential observational biases, and improve confidence in automated classification systems (Mohammadi et al., 2022; Kumaran et al., 2025; Mbouopda et al., 2026). Nevertheless, explainability alone cannot establish physical causality. Interpretable predictions must still be evaluated against established astrophysical theory and independent observational evidence. The emergence of scientific foundation models further expands the scope of Hybrid Scientific AI. Inspired by the success of large foundation models in natural language processing and computer vision, these architectures aim to learn general representations from heterogeneous scientific datasets that can subsequently be adapted to multiple downstream tasks. In astronomy, foundation models could integrate images, spectra, time-series observations, simulation outputs, and textual scientific knowledge within a common representation, enabling transfer learning across different astronomical domains while reducing dependence on task-specific training data. An even more ambitious vision is provided by digital twins, which combine observational data assimilation, numerical simulations, uncertainty quantification, and continuous machine learning updates to create dynamic digital representations of physical systems. Rather than producing static predictions, digital twins evolve continuously as new observations become available, providing a framework for real-time monitoring, hypothesis testing, and predictive modelling. Recent reviews have highlighted the potential of combining digital twins with Bayesian learning, generative AI, and large language models to support autonomous scientific workflows (Sun et al., 2025; Zhou et al., 2026). Although applications within astronomy remain in their infancy, similar concepts could eventually support adaptive telescope scheduling, real-time space-weather forecasting, autonomous observatories, and continuously updated cosmological simulations. Despite its considerable promise, Hybrid Scientific AI also introduces new scientific and computational challenges. Integrating heterogeneous observational datasets, numerical simulations, and multiple machine learning components substantially increases model complexity. Ensuring reproducibility, validating predictions against physical theory, quantifying uncertainty, and maintaining computational efficiency remain open research problems. In addition, many existing hybrid frameworks have been demonstrated only within controlled benchmark studies, and their scalability to operational astronomical environments has yet to be established. Addressing these challenges will require closer collaboration between astronomers, computational physicists, machine learning researchers, and high-performance computing specialists. From a broader perspective, Hybrid Scientific AI represents the current frontier of computational astronomy. Rather than replacing either physics-based modelling or conventional machine learning, it establishes a collaborative framework in which observations, simulations, physical theory, and artificial intelligence reinforce one another. As astronomical datasets continue to expand in both size and complexity, this integrated approach is likely to become essential for extracting reliable scientific knowledge from next-generation observational facilities. A comparative summary of the major artificial intelligence paradigms discussed in this review is presented in Table 2, highlighting their level of physics integration, scientific strengths, limitations, and suitability for different astronomical applications. Critical Perspective. Hybrid Scientific AI marks the transition from predictive machine learning to computational systems capable of supporting scientific discovery. Its greatest strength lies in combining complementary sources of information observations, simulations, governing physics, uncertainty estimation, and explainability within a unified framework. Although significant methodological and computational challenges remain, this paradigm provides the most promising foundation for developing trustworthy, interpretable, and physically consistent AI systems capable of addressing the increasingly complex scientific questions of modern astronomy and astrophysics.

Table 2. Critical comparison of artificial intelligence paradigms in astronomy and astrophysics

Methodology	Physics Integration	Data Requirement	Computational Cost	Interpretability	Generalisation Beyond Training Data	Current Maturity	Representative Applications	Major Strength	Principal Limitation
Pure Machine Learning	Low	High	Low	Low	Limited	Mature	Galaxy classification, stellar parameter estimation, transient detection	Fast training and high predictive performance	Predictions may violate physical laws and perform poorly outside the training distribution
Physics-Guided Learning	Moderate	Moderate	Low–Moderate	Moderate	Moderate	Mature	Feature-based stellar modelling, cosmological parameter estimation	Incorporates domain knowledge with minimal computational overhead	Physical consistency depends on engineered features rather than optimisation
Physics-Constrained Learning	Moderate–High	Moderate	Moderate	Moderate	Improved	Developing	Hybrid simulation–data models, constrained optimisation	Improves robustness while preserving model flexibility	Requires careful design and balancing of physical constraints
Physics-Informed Neural Networks (PINNs)	High	Low–Moderate	High	High	High	Mature	Solar magnetic fields, astrophysical shocks, neutron stars, inverse problems	Governing equations are directly embedded into learning	High computational cost, optimisation instability and scalability challenges
Neural Operators	High	High (simulation-based)	High (training), Very Low (inference)	Moderate	High	Emerging	Surrogate modelling, cosmological simulations, uncertainty analysis	Rapid prediction across multiple physical configurations	Dependence on large, high-quality simulation datasets
Hybrid Scientific AI	Very High	Multi-source	High	High	Very High	Emerging	Scientific workflows, multi-fidelity modelling, uncertainty-aware AI	Combines observations, simulations and physical reasoning	Increased implementation complexity and validation requirements
Digital Twins & Autonomous Scientific AI	Very High	Continuous observational and simulation data	Very High	High	Very High	Early Emerging	Autonomous observatories, adaptive simulations, real-time space weather forecasting	Continuous learning with real-time model updating	Limited astronomical deployment and high computational infrastructure requirements

The evolution of artificial intelligence in astronomy reflects a clear transition from purely data-driven prediction towards physically informed and scientifically robust learning. Conventional machine learning remains indispensable for analysing large observational datasets because of its computational efficiency and strong predictive performance. However, its reliance on statistical correlations alone limits its ability to produce physically consistent and reliable predictions when applied beyond the training domain. Physics-guided and physics-constrained approaches address many of these limitations by incorporating domain knowledge into the learning process, providing a more effective balance between computational efficiency and physical realism. Physics-Informed Neural Networks (PINNs) represent a further step forward by embedding governing equations directly within model optimisation, making them particularly well suited to inverse problems and data-sparse astrophysical applications. Nevertheless, their high computational cost and optimisation challenges continue to restrict their routine application to large-scale simulations. Neural operators alleviate many of these limitations by learning solution operators rather than individual solutions, enabling rapid surrogate modelling of complex physical systems. Building on these advances, Hybrid Scientific AI combines observational data, numerical simulations, uncertainty quantification, explainable artificial intelligence, and emerging technologies such as scientific foundation models and digital twins within unified computational frameworks. Collectively, these developments redefine artificial intelligence from a purely predictive tool into an integrated component of the scientific discovery process, capable of supporting hypothesis generation, accelerating simulations, and improving scientific interpretation. Overall, no single methodology is universally applicable. The most appropriate approach depends on the scientific objective, the availability and quality of observational or simulated data, the complexity of the governing physical processes, and the level of scientific reliability required. Rather than replacing conventional machine learning, increasing levels of physics integration expand its capability, allowing models to become more robust, interpretable, and physically consistent. This progression suggests that future advances in astronomical artificial intelligence will depend not only on improvements in machine learning algorithms but also on closer integration with physical theory, numerical modelling, and uncertainty-aware scientific reasoning. Recent developments further demonstrate that the impact of physics-informed machine learning extends beyond methodological innovation alone. Explainable artificial intelligence is increasingly being explored to improve astronomy education and data literacy, helping researchers and students better interpret complex astronomical datasets while promoting more transparent use of AI tools (Marín Díaz, 2025). At the same time, practical tutorial studies have made physics-informed neural networks more accessible by providing benchmark implementations for astrophysical and plasma physics applications, thereby lowering the barrier to their adoption within the broader scientific community (Baty, 2024). Together, these advances highlight that sustained progress in physics-informed machine learning will rely not only on new algorithms but also on reproducible computational frameworks, accessible educational resources, and the effective dissemination of methodological knowledge.

4. Critical Assessment Across Astronomy

The rapid growth of physics-informed machine learning (PIML) has reshaped the role of artificial intelligence in astronomy, shifting its purpose from automated prediction towards scientifically meaningful inference. While conventional machine learning has substantially improved the speed and efficiency of astronomical data analysis, many modern applications require models that are physically consistent, interpretable, and capable of generalising beyond the conditions represented in the training data. Integrating physical laws, numerical simulations, and observational evidence has therefore become increasingly important for developing trustworthy AI systems capable of supporting scientific discovery (Hao et al., 2022; Meng et al., 2025). Despite this progress, the maturity of PIML varies considerably across different branches of astronomy. Areas supported by well-established physical models, such as solar physics and computational fluid dynamics, have demonstrated clear benefits from incorporating physical constraints (Jarolim et al., 2023; Moschou et al., 2023). In contrast, applications relying on sparse observations or highly uncertain physical processes continue to depend largely on conventional data-driven learning. Consequently, evaluating machine learning solely on predictive accuracy is no longer sufficient. A more meaningful assessment considers whether these methods improve physical understanding, produce reliable uncertainty estimates, and contribute to scientific interpretation. The following sections critically examine these aspects across major astronomical research domains.

4.1 Stellar Evolution

Stellar evolution has become one of the most successful applications of scientific machine learning because modern surveys, including Gaia, Kepler, and TESS, produce vast quantities of photometric, spectroscopic, and asteroseismic data. Machine learning has dramatically accelerated the estimation of stellar mass, age, radius, effective temperature, luminosity, and internal structural properties, reducing computations that traditionally required iterative stellar evolution modelling from hours to seconds (Bellinger et al., 2016; Mombarg et al., 2021; Li et al., 2025). The principal strength of these approaches lies in computational efficiency. Surrogate models enable rapid population synthesis, Bayesian parameter estimation, and uncertainty analysis across millions of stars, making large-scale stellar surveys computationally feasible. Deep learning has also improved the interpretation of asteroseismic observations by identifying complex relationships that are difficult to capture using conventional optimisation methods (Mombarg et al., 2021). However, most existing models remain strongly dependent on the quality and diversity of their training data. Predictions become less reliable for stellar populations with unusual metallicities, rapid rotation, strong magnetic activity, or binary interactions because these systems are relatively underrepresented in available observational catalogues. More importantly, many machine learning models reproduce stellar observables without explicitly modelling the underlying physical processes, including convection, rotational mixing, mass loss, turbulence, and magnetic evolution. Consequently, high predictive accuracy does not necessarily translate into improved physical understanding. Future progress will require tighter integration of stellar structure equations, evolutionary constraints, uncertainty quantification, and multi-fidelity simulation data within machine learning frameworks. Rather than serving solely as fast surrogate models, future systems should support the refinement of stellar evolution theory itself by combining observational evidence with physically consistent learning.

4.2 Exoplanets

Machine learning has become an important component of modern exoplanet research, supporting transit detection, candidate validation, orbital parameter estimation, atmospheric retrieval, and habitability assessment. Deep learning models have enabled efficient analysis of the vast photometric datasets produced by missions such as Kepler, TESS, and the upcoming PLATO mission, substantially reducing the need for manual inspection and accelerating the discovery of planetary candidates (Meher & Panda, 2021; Ting, 2025). One of the greatest strengths of these approaches is their ability to detect weak transit signals embedded within noisy stellar light curves. However, their performance remains strongly dependent on the quality and representativeness of the training data. Existing exoplanet catalogues are affected by significant observational selection biases because current detection techniques preferentially identify large, short-period planets. As a result, machine learning models may learn these observational biases rather than the intrinsic diversity of planetary systems, limiting their ability to generalise to Earth-sized planets, long-period orbits, or poorly sampled stellar populations. Current explainable AI methods improve model transparency by identifying influential observational features, but they do not establish whether these features are physically related to planetary formation, atmospheric evolution, or orbital dynamics. Feature importance therefore provides useful insight into model behaviour but should not be interpreted as evidence of astrophysical causality (Mohammadi et al., 2022; Kumaran et al., 2025). Future progress is likely to depend on integrating data-driven learning with established physical knowledge. Incorporating orbital dynamics, atmospheric chemistry, radiative transfer, and planetary evolution models can reduce physically implausible predictions, improve extrapolation to underrepresented planetary systems, and provide more scientifically meaningful interpretations. Consequently, conventional machine learning remains highly effective for large-scale candidate screening, whereas physics-informed approaches are likely to play an increasingly important role in characterising planetary systems and assessing the habitability of worlds beyond the Solar System.

4.3 Galaxy Formation and Evolution

Galaxy formation remains one of the most computationally demanding problems in astrophysics because it involves the interaction of gravity, hydrodynamics, star formation, chemical enrichment, feedback processes, and dark matter across multiple spatial and temporal scales. Machine learning has increasingly been adopted to accelerate cosmological simulations, estimate galaxy properties, classify morphologies, and develop surrogate models that emulate computationally expensive numerical calculations (Villaescusa-Navarro et al., 2021; Kugel et al., 2023). The most significant contribution of machine learning in this area is computational acceleration. Modern surrogate models can reproduce selected outputs of large cosmological simulations at a fraction of the computational cost, enabling extensive parameter exploration, Bayesian inference, and uncertainty analysis that would otherwise require thousands of expensive numerical simulations. Despite these advances, important scientific challenges remain. Many machine learning models are trained using data generated by a single simulation framework and consequently inherit its underlying assumptions, calibration strategies, and numerical approximations. Agreement with simulated data therefore should not be interpreted as direct validation against the physical Universe. Furthermore, models trained on one cosmological simulation frequently exhibit reduced performance when transferred to alternative simulations or observational datasets, indicating that many learned relationships remain simulation-dependent rather than universally physical. Future research should prioritise cross-simulation generalisation, observational validation, uncertainty-aware surrogate modelling, and hybrid frameworks that integrate observational evidence with simulation-derived knowledge. These developments will be essential for ensuring that machine learning contributes not only to faster computation but also to more reliable cosmological inference.

4.4 Gravitational-Wave Astronomy

The detection of gravitational waves has created new computational challenges because waveform analysis, parameter estimation, and source localisation must often be performed within minutes of an event. Machine learning has therefore become an important component of modern gravitational-wave astronomy, enabling rapid event detection, waveform reconstruction, and astrophysical source characterisation. Compared with conventional Bayesian inference, surrogate models substantially reduce computational time while maintaining competitive predictive accuracy, making them particularly valuable for real-time multi-messenger observations. Despite these advances, scientific reliability remains a major concern. Gravitational-wave signals are frequently characterised by low signal-to-noise ratios, detector artefacts, and rare astrophysical events that are poorly represented in available training datasets. Consequently, deterministic predictions alone are insufficient for robust scientific inference. Reliable uncertainty quantification and consistency with general relativity are essential because even small modelling errors may influence estimates of source masses, spins, and merger dynamics.

Physics-informed machine learning provides a promising solution by integrating waveform physics with observational data, thereby reducing physically unrealistic predictions while preserving computational efficiency. Future research should focus on combining Bayesian inference, uncertainty-aware learning, and physics-constrained optimisation within unified frameworks capable of supporting both rapid and scientifically reliable gravitational-wave analysis.

4.5 Space Weather

Space-weather forecasting represents one of the most operationally important applications of scientific AI because accurate prediction of solar flares, coronal mass ejections, and geomagnetic storms directly affects satellite operations, navigation systems, communication networks, and electrical power infrastructure. The continuous availability of solar observations has enabled machine learning models to identify complex patterns associated with solar activity and improve forecasting performance. Recent advances in Physics-Informed Neural Networks have demonstrated that incorporating magnetohydrodynamic constraints improves the reconstruction of solar magnetic fields and produces physically consistent predictions compared with purely statistical approaches (Jarolim et al., 2023). Nevertheless, many operational forecasting systems remain predominantly data-driven, relying on historical correlations rather than explicit representations of the underlying physical processes governing solar activity. Operational deployment introduces additional challenges beyond predictive accuracy. Forecasting systems must operate continuously under changing observational conditions while providing robust uncertainty estimates, computational efficiency, and interpretable predictions suitable for operational decision-making. Future developments are therefore expected to integrate magnetohydrodynamic models, real-time data assimilation, explainable AI, and uncertainty quantification within adaptive forecasting systems capable of continuously updating their predictions as new observations become available. To provide a domain-specific perspective, Table 3 presents a critical assessment of physics-informed machine learning across the major areas of astronomy and astrophysics, highlighting the current level of methodological maturity, scientific strengths, and remaining research challenges.

The current literature demonstrates that physics-informed machine learning has progressed beyond simply improving predictive accuracy. Across astronomy, the greatest advances have been achieved where well-established physical models can be integrated directly into the learning process, enabling more robust predictions from sparse or noisy observations. However, the degree of scientific maturity differs substantially between research domains. Stellar evolution and space-weather forecasting have benefited considerably from the incorporation of physical constraints, whereas galaxy formation and gravitational-wave astronomy continue to face challenges associated with computational complexity, uncertainty quantification, and validation against independent observations. A common limitation across nearly all application areas is the continued reliance on benchmark datasets or numerical simulations. Although these resources are essential for model development, agreement with simulated data does not necessarily guarantee agreement with physical reality. Consequently, future research should place greater emphasis on observational validation, cross-survey generalisation, uncertainty-aware prediction, and reproducible benchmarking. These developments will be critical for establishing artificial intelligence not merely as a computational accelerator but as a scientifically reliable framework capable of supporting hypothesis generation, observational planning, and physical discovery. Overall, the evidence suggests that the future of astronomical AI lies not in replacing physics with machine learning, but in combining both within unified scientific frameworks. Hybrid approaches that integrate observations, governing equations, numerical simulations, uncertainty quantification, and explainable AI provide the most promising pathway towards trustworthy Scientific AI capable of addressing the increasingly complex challenges posed by next-generation astronomical facilities. Fig. 4 provides an overview of the current landscape of physics-informed machine learning in astronomy and astrophysics, summarising its principal application areas and emerging research directions.

Table 3. Critical assessment of physics-informed machine learning across major astronomical domains

Astronomy Domain	Current Strength	Major Limitation	Scientific Maturity	Future Research Priority
Stellar Evolution	Rapid estimation of stellar parameters from large surveys	Limited representation of rare stellar populations and incomplete modelling of stellar physics	High	Multi-fidelity learning integrating stellar evolution theory with observational data
Exoplanets	Efficient transit detection and candidate classification	Observational bias, limited interpretability, and uncertainty estimation	High	Physics-aware planetary formation and atmospheric models
Galaxy Formation	Fast surrogate modelling of cosmological simulations	Simulation-dependent learning and limited cross-simulation generalisation	Moderate–High	Observation-constrained hybrid simulation–AI frameworks
Gravitational-Wave Astronomy	Rapid waveform reconstruction and source parameter estimation	Limited uncertainty quantification and dependence on simulated training data	Moderate	Physics-informed Bayesian inference and uncertainty-aware learning
Space Weather	Real-time forecasting of solar activity	Limited integration of magnetohydrodynamic physics into operational systems	High	Digital twins with continuous observational assimilation

The conceptual bubble chart illustrates the relative maturity, level of physics integration, and machine learning adoption across major application domains. Bubble size represents research activity, the horizontal axis represents machine learning sophistication, and the vertical axis represents the degree of physical knowledge integration. Stellar evolution and space weather currently exhibit strong integration between physical modelling and machine learning, whereas exoplanet research remains comparatively data-driven. Galaxy formation and gravitational-wave astronomy occupy intermediate positions due to their growing reliance on simulation-assisted and physics-constrained learning frameworks.

5. Scientific Reliability: Are Physics-Informed Models Really Better?

Physics-informed machine learning (PIML) has emerged as a promising alternative to conventional data-driven artificial intelligence by integrating physical laws with machine learning. This integration is expected to improve scientific reliability, particularly in astronomy where observations are sparse, noisy, and often impossible to reproduce experimentally. However, whether PIML consistently outperforms conventional machine learning remains an open question. Current evidence indicates that no single modelling paradigm is universally superior. Instead, model performance depends on the availability of observational data, the complexity of the governing physics, computational resources, and the scientific objective of the study (Hao et al., 2022; Meng et al., 2025). Rather than evaluating models solely through prediction accuracy, scientific reliability should be assessed using multiple complementary criteria, including physical consistency, robustness to distribution shifts, computational efficiency, interpretability, uncertainty quantification, and reproducibility. These criteria provide a more realistic measure of whether AI models can support scientific discovery rather than simply generating accurate predictions.

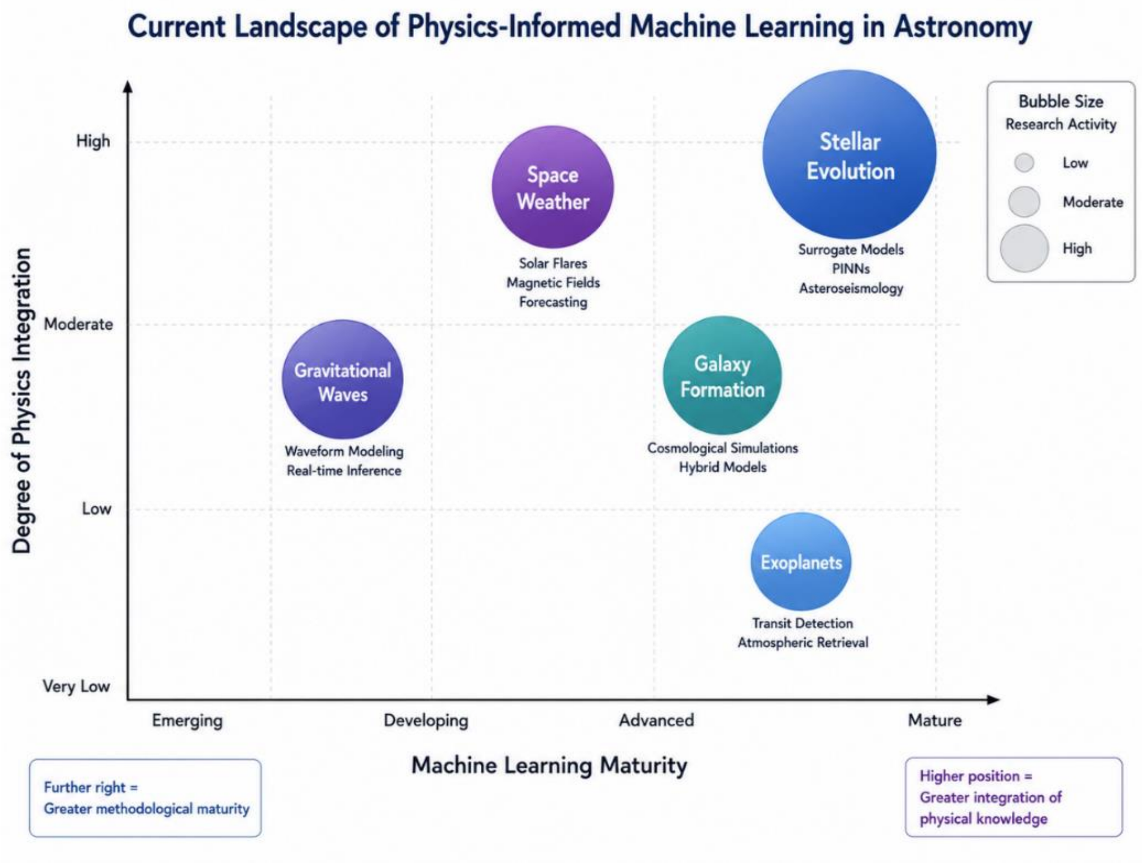

Fig. 4. Current Landscape of Physics-Informed Astronomy Current landscape of physics-informed machine learning applications in astronomy and astrophysics

5.1 Predictive Accuracy: Does Physics Always Improve Performance?

A common assumption is that incorporating physical knowledge automatically improves predictive accuracy. Current evidence suggests that this is not universally true. For interpolation tasks supported by large, high-quality observational datasets, conventional machine learning often performs exceptionally well because it can efficiently learn complex nonlinear relationships directly from the data (Ball & Brunner, 2010; Ting, 2025). In these situations, introducing additional physical constraints may increase optimisation complexity while offering only modest gains in prediction accuracy. The advantages of PIML become more apparent when observations are sparse, noisy, or costly to obtain. Physical constraints reduce the range of admissible solutions, helping models avoid physically unrealistic predictions and improving stability under limited data. Consequently, the primary strength of PIML lies not in consistently achieving higher accuracy, but in producing predictions that remain physically plausible under challenging observational conditions (Hao et al., 2022; Ghalambaz et al., 2024).

5.2 Generalisation Beyond the Training Domain

Generalisation remains one of the most demanding challenges in astronomical machine learning because many scientific investigations involve rare objects, extreme environments, or previously unobserved physical conditions. Conventional machine learning models typically perform well within the distribution represented by the training data but often lose accuracy when applied to different surveys or unseen parameter ranges. By embedding governing equations or physical constraints, PIML reduces the likelihood of physically implausible extrapolations and generally provides greater robustness under distribution shifts. However, improved extrapolation is not guaranteed. Model performance still depends on the validity and completeness of the incorporated physical knowledge. Simplified assumptions or incomplete physical descriptions can introduce systematic errors even when physical constraints are formally satisfied (Meng et al., 2025). Physics therefore strengthens the potential for generalisation but does not remove the need for rigorous validation using independent observational data.

5.3 Computational Efficiency

Improved physical realism is often accompanied by increased computational cost. Conventional machine learning models optimise statistical loss functions and can usually be trained efficiently on large datasets, making them well suited for large-scale astronomical surveys and real-time data analysis. In contrast, physics-informed approaches require repeated evaluation of governing equations, conservation laws, or residual functions during optimisation, substantially increasing training time and memory requirements compared with purely data-driven models (Hao et al., 2022). Recent advances, including neural operators and physics-informed surrogate models, seek to overcome these limitations by learning reusable solution operators from numerical simulations. Although these methods require an expensive offline training stage, they can rapidly predict solutions for new physical configurations without repeatedly solving the underlying governing equations, making them particularly attractive for uncertainty quantification, parameter exploration, and large-scale astrophysical simulations (Lu & Wang, 2024). These observations highlight an important trade-off between computational efficiency and physical fidelity. Conventional machine learning offers rapid training and inference but relies primarily on statistical correlations. Physics-informed methods provide stronger physical consistency but at a substantially higher computational cost, particularly for high-dimensional and multi-physics problems. Neural operators attempt to bridge this gap by combining fast inference with physics-aware representations learned from simulation data. Improving computational scalability while preserving physical realism therefore remains one of the key challenges in developing the next generation of trustworthy scientific AI for astronomy and astrophysics.

5.4 Physical Consistency

The most significant advantage of PIML is its ability to improve physical consistency. Conventional machine learning evaluates success primarily through statistical error metrics and may therefore produce predictions that violate conservation laws or governing equations. PIML reduces this risk by embedding physical principles directly into the optimisation process, resulting in predictions that are generally more consistent with established scientific theory (Jarolim et al., 2023; Dai et al., 2024). However, satisfying physical constraints should not be interpreted as proof of scientific correctness. If the governing equations are simplified or important physical processes are omitted, the resulting predictions may remain systematically biased. Consequently, the reliability of PIML depends as much on the quality of the embedded physics as on the learning algorithm itself.

5.5 Interpretability and Trust

Interpretability is another frequently cited advantage of PIML. Conventional deep learning models often behave as black boxes, making it difficult to determine whether predictions arise from meaningful physical relationships or spurious statistical correlations. By incorporating physical knowledge, PIML provides a more transparent connection between model behaviour and established scientific principles (Pilania et al., 2018; Singh et al., 2023). Nevertheless, interpretability remains relative rather than absolute. Large physics-informed neural networks still contain millions of trainable parameters, and explainable AI techniques such as SHAP or LIME identify influential variables rather than causal physical mechanisms. Scientific interpretation therefore continues to require independent validation against astrophysical theory and observations (Kumaran et al., 2025).

5.6 Towards Trustworthy Scientific AI

The ultimate objective of scientific AI is not simply accurate prediction but reliable scientific inference. Future evaluation frameworks should therefore extend beyond conventional metrics such as accuracy, root mean square error, or coefficient of determination. Robust validation should include agreement with established physical theory, uncertainty quantification, reproducibility across independent datasets, performance under distribution shifts, computational efficiency, and physical interpretability (Fluke & Jacobs, 2020; Meng et al., 2025). Critical Perspective. Current evidence suggests that physics-informed machine learning should not be viewed as a universal replacement for conventional machine learning. Data-driven models remain highly effective for well-sampled observational problems, whereas PIML offers its greatest benefits in data-sparse, physics-rich applications where physical consistency and extrapolation are essential. The future of astronomy is therefore unlikely to be defined by competition between machine learning and physics. Instead, scientific progress will depend on integrating both within trustworthy AI frameworks that combine predictive accuracy with physical reasoning, uncertainty awareness, and rigorous scientific validation. The progressive improvement in scientific reliability with increasing levels of physics integration is conceptually illustrated in Fig. 5, highlighting the transition from conventional data-driven machine learning to physics-informed and scientific AI methodologies.

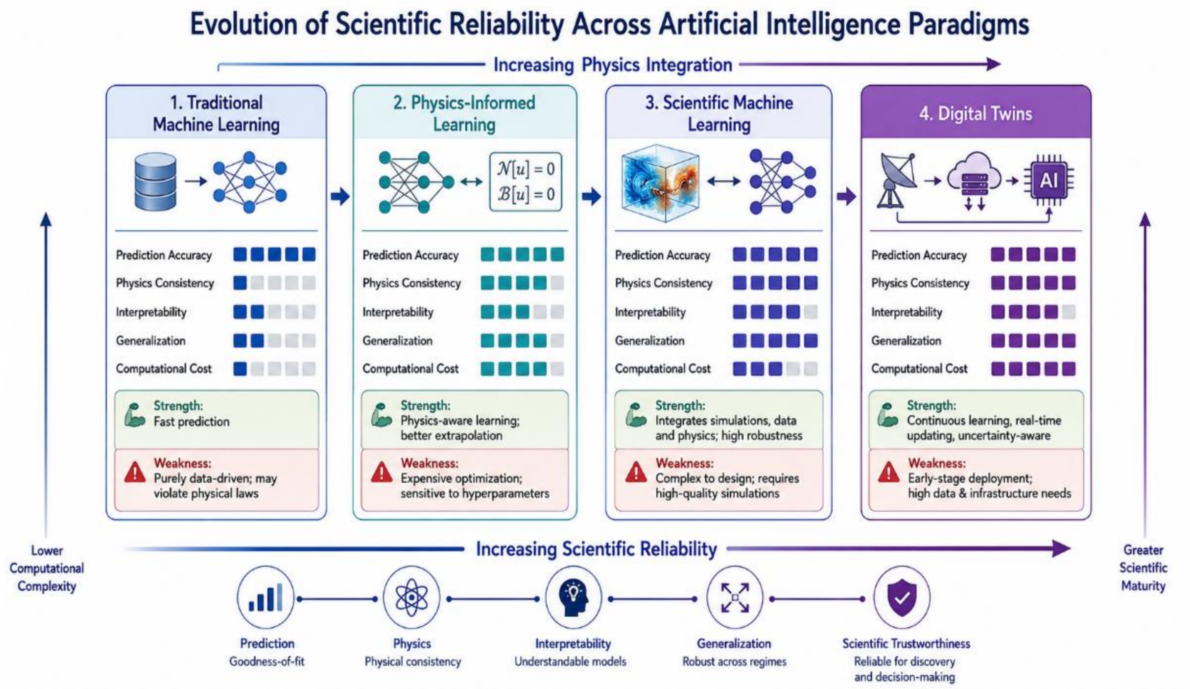

Fig. 5. Evolution of Scientific Reliability Across Artificial Intelligence Paradigms

Comparison of the progression from traditional machine learning to physics-informed learning, scientific machine learning, and digital twins. The conceptual framework illustrates how increasing integration of physical knowledge generally improves physical consistency, scientific reliability, and generalization, while simultaneously increasing computational complexity and implementation requirements. The figure emphasizes that methodological sophistication should be evaluated using multiple criteria rather than predictive accuracy alone. Table 4 provides a critical comparison of the major artificial intelligence paradigms in astronomy and astrophysics, highlighting their scientific strengths, computational characteristics, and principal limitations.

The comparison demonstrates that no single AI paradigm is universally optimal. Conventional machine learning remains the preferred choice when large, representative observational datasets are available and rapid prediction is the primary objective. As observational data become sparse, noisy, or extend beyond the training distribution, the value of incorporating physical knowledge increases substantially. Physics-guided and physics-constrained approaches provide an effective compromise between computational efficiency and physical realism, whereas PINNs are particularly advantageous for inverse problems governed by well-established differential equations. Neural operators address many of the computational limitations of PINNs by learning reusable solution operators, making them attractive for accelerating large-scale numerical simulations. Looking further ahead, Hybrid Scientific Machine Learning and digital twins represent the emerging direction of the field by combining observations, simulations, governing physics, uncertainty quantification, and explainable AI within unified computational frameworks. Overall, the evidence indicates that the future of trustworthy astronomical AI will depend not on replacing conventional machine learning but on selecting the most appropriate level of physics integration for a given scientific problem. Consequently, scientific reliability should be evaluated using a combination of predictive accuracy, physical consistency, uncertainty quantification, computational scalability, reproducibility, and interpretability rather than conventional statistical metrics alone.

6. Explainability, Uncertainty and Trustworthy Scientific AI

As artificial intelligence becomes increasingly embedded in astronomical research, evaluating models solely by predictive accuracy is no longer sufficient. Scientific applications require AI systems that not only produce reliable predictions but also provide interpretable reasoning, quantify uncertainty, and support reproducible research. These characteristics are essential for ensuring that machine learning contributes to scientific understanding rather than functioning as a black-box prediction tool (Fluke & Jacobs, 2020; Hao et al., 2022; Meng et al., 2025). Trustworthy scientific AI therefore extends beyond algorithmic performance to encompass explainability, uncertainty quantification, reproducibility, transparent benchmarking, and open science. The following sections critically examine how these complementary principles contribute to the development of reliable and scientifically credible AI for astronomy and astrophysics.

6.1 Explainability Beyond Feature Importance

The remarkable success of deep learning has also highlighted one of its greatest weaknesses—limited interpretability. Modern neural networks often contain millions of trainable parameters, making it difficult to determine why a particular prediction has been generated. This lack of transparency is particularly problematic in astronomy, where AI is increasingly used to support scientific interpretation rather than routine data processing (Baron, 2019; Ting, 2025). Post hoc explainability techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) have become widely adopted because they identify the relative contribution of input variables to individual predictions. These methods have been successfully applied to astronomical source classification, stellar parameter estimation, galaxy morphology analysis, and transient-event identification, providing researchers with greater insight into model behaviour (Mohammadi et al., 2022; Kumaran et al., 2025; Mbouopda et al., 2026). However, feature importance should not be confused with physical understanding. Variables identified as influential may simply reflect statistical patterns or observational biases present in the training data rather than genuine astrophysical mechanisms. Consequently, explainability alone cannot establish causality or validate scientific hypotheses. A more meaningful objective is physics-aware explainability, where model explanations are evaluated against established astrophysical theory. Instead of asking which variable contributes most to a prediction, the emphasis shifts towards determining whether the model's reasoning is consistent with known physical processes. This transition moves explainability beyond visual interpretation towards scientifically meaningful validation, strengthening confidence in AI-assisted discoveries.

6.2 Uncertainty Quantification

Scientific conclusions require more than point predictions; they also require reliable estimates of confidence. Astronomical observations are inherently uncertain because they are affected by instrumental noise, incomplete sampling, observational bias, and intrinsic variability. Ignoring these uncertainties may lead to overconfident predictions and unreliable scientific conclusions. Conventional deterministic neural networks typically provide a single prediction without indicating its reliability. This limitation becomes particularly important when models are applied to rare astrophysical events or conditions outside the training distribution. In contrast, probabilistic approaches including Bayesian neural networks, Gaussian processes, Monte Carlo dropout, and ensemble learning estimate predictive uncertainty alongside model outputs, enabling researchers to distinguish between well-supported predictions and those requiring further investigation (Hao et al., 2022; Meng et al., 2025). Rather than representing a weakness, uncertainty should be viewed as valuable scientific information. Models that communicate their confidence allow astronomers to prioritise follow-up observations, assess prediction reliability, and reduce the risk of drawing conclusions from uncertain data. As astronomy increasingly relies on automated decision-making, uncertainty-aware AI will become an essential component of trustworthy scientific workflows.

Table 4. Critical comparison of artificial intelligence paradigms for scientifically reliable astronomy

AI Paradigm	Predictive Performance	Physical Fidelity	Computational Demand	Scientific Interpretability	Extrapolation Capability	Most Suitable Applications	Principal Limitation
Traditional Machine Learning	High for interpolation tasks	Low	Low	Low–Moderate	Limited	Large observational datasets, classification, regression, anomaly detection	Learns statistical correlations rather than physical mechanisms, leading to unreliable predictions outside the training domain
Physics-Guided Learning	High	Moderate	Low–Moderate	Moderate	Moderate	Feature-based prediction, stellar parameter estimation, cosmological inference	Physical knowledge influences feature design but does not constrain optimisation, allowing physically inconsistent solutions
Physics-Constrained Learning	High	Moderate–High	Moderate	Moderate–High	Good	Hybrid simulation–data modelling, inverse problems	Performance depends strongly on selecting appropriate physical constraints and balancing optimisation objectives
Physics-Informed Neural Networks (PINNs)	High for sparse-data and inverse problems	High	High	High	High	Solar physics, astrophysical fluid dynamics, neutron stars, gravitational physics	High computational cost, optimisation instability, and limited scalability for complex multi-physics systems
Neural Operators	High	High	High (training), Low (inference)	Moderate	High	Large-scale surrogate modelling, cosmological simulations, uncertainty analysis	Dependence on extensive high-quality simulation datasets and limited physical interpretability
Hybrid Scientific Machine Learning	High	High	Moderate–High	High	High	Multi-fidelity learning, observation–simulation integration, uncertainty-aware inference	Requires careful integration of heterogeneous datasets, physical models, and validation strategies
Digital Twins and Autonomous Scientific AI	Potentially Very High	Very High	Very High	High	Very High	Real-time monitoring, adaptive simulations, autonomous observatories, operational space weather	Early-stage development requiring continuous observational data, high-performance computing, and rigorous validation

6.3 Scientific Reproducibility

Reproducibility is a cornerstone of scientific research, yet achieving it remains more challenging for machine learning than for conventional numerical simulations. Variations in data preprocessing, random initialisation, hyperparameter optimisation, software libraries, and computing hardware can all influence model performance, making direct comparison between published studies difficult and sometimes limiting confidence in reported results (Fluke & Jacobs, 2020). Improving reproducibility requires more than reporting predictive accuracy. Studies should provide transparent documentation of datasets, preprocessing pipelines, model architectures, training protocols, hyperparameters, evaluation metrics, and computational environments. Public release of source code, trained models, benchmark datasets, and reproducible workflows further enables independent verification and facilitates objective comparison across competing methodologies. For astronomy, reproducibility has particular importance because AI models are increasingly used to support scientific interpretation rather than simply automate data processing. Results that cannot be independently reproduced are difficult to validate and have limited value for advancing scientific understanding. Consequently, reproducibility should be regarded as a fundamental requirement for trustworthy scientific AI rather than a desirable reporting practice. Wider adoption of open-source software, standardised benchmark datasets, and transparent reporting guidelines will be essential for ensuring that future AI models are reliable, comparable, and suitable for scientific discovery.

6.4 Benchmarking and Standardised Evaluation

Despite rapid advances in astronomical machine learning, objective comparison between published models remains difficult because studies often use different datasets, preprocessing procedures, train–test splits, and evaluation metrics. Consequently, improvements reported in one study cannot always be compared directly with those reported elsewhere, limiting the reproducibility and practical adoption of new methodologies (Fluke & Jacobs, 2020; Sen et al., 2022). Future benchmarking should extend beyond conventional statistical metrics such as accuracy, mean squared error, or the coefficient of determination. A scientifically meaningful evaluation framework should also assess physical consistency, uncertainty calibration, computational efficiency, robustness under distribution shifts, interpretability, and reproducibility across independent datasets (Hao et al., 2022; Meng et al., 2025). Such multidimensional benchmarking would provide a more balanced assessment of model quality and encourage the development of AI systems that are both computationally effective and scientifically reliable.

6.5 Open Science and Trustworthy AI

The long-term success of scientific AI depends on transparency and reproducibility as much as algorithmic innovation. Open access to observational datasets, numerical simulations, benchmark repositories, trained models, and source code enables independent validation and accelerates methodological progress across the astronomical community. Equally important is the transparent reporting of model assumptions, limitations, uncertainty estimates, and domains of applicability, allowing researchers to understand when and where a model can be used reliably. As AI increasingly contributes to hypothesis generation, observational planning, and scientific interpretation, trust cannot be established through predictive performance alone. Instead, trustworthy scientific AI requires the integration of explainability, uncertainty quantification, reproducible workflows, rigorous benchmarking, and open scientific practices. These principles ensure that AI models remain transparent, verifiable, and scientifically credible, facilitating their adoption as reliable research tools rather than isolated predictive algorithms. To synthesize the key characteristics of approaches that contribute to scientifically trustworthy artificial intelligence, Table 5 compares the major methodologies in terms of their scientific reliability, interpretability, physical consistency, uncertainty awareness, and practical applicability.

The future of scientific AI depends on far more than achieving high predictive accuracy. Explainability should provide insight into physically meaningful reasoning rather than simply ranking influential features. Uncertainty quantification should accompany predictions to support informed scientific decision-making, particularly under sparse observations or previously unseen conditions. Reproducible workflows, standardised benchmarking, and open science practices are equally essential for ensuring that published models can be independently validated and fairly compared. Taken together, these principles define the foundations of trustworthy scientific AI. Future progress in astronomy will depend not only on developing more powerful algorithms but also on ensuring that these algorithms are transparent, reproducible, uncertainty-aware, and consistent with established physical knowledge. Such systems will provide greater confidence in AI-assisted discoveries and enable artificial intelligence to become a reliable partner in scientific exploration rather than simply a high-performance prediction tool (Fluke & Jacobs, 2020; Hao et al., 2022; Meng et al., 2025).

Table 5. Critical comparison of approaches supporting trustworthy scientific AI

Technique	Primary Objective	Key Strength	Major Limitation	Contribution to Scientific Reliability
SHAP	Global and local feature attribution	Provides intuitive interpretation of influential variables	Explains statistical associations rather than physical causality	Improves transparency and supports model inspection
LIME	Local explanation of individual predictions	Model-agnostic and applicable to diverse algorithms	Explanations may vary across repeated analyses and complex models	Assists interpretation of individual predictions
Bayesian Machine Learning	Predictive uncertainty estimation	Produces probabilistic predictions and confidence intervals	Increased computational cost and implementation complexity	Enables uncertainty-aware scientific inference
Ensemble Learning	Robustness assessment	Reduces prediction variance and improves model stability	Requires multiple models and greater computational resources	Increases confidence and robustness
Physics-Aware Explainability	Interpret predictions using physical knowledge	Connects AI decisions with established scientific principles	Requires validated physical models and domain expertise	Bridges statistical learning with astrophysical reasoning
Benchmarking Frameworks	Standardised model evaluation	Enables objective comparison across studies	Requires common datasets and agreed evaluation protocols	Improves reproducibility and methodological consistency
Open Science Practices	Transparency and reproducibility	Facilitates independent validation, collaboration, and reuse	Additional effort for documentation, maintenance, and data sharing	Strengthens long-term scientific credibility and community trust

7. Future Research Directions: Towards Autonomous Scientific Artificial Intelligence

Artificial intelligence in astronomy is undergoing a fundamental transition from data-driven prediction towards scientifically informed reasoning. Early machine learning methods primarily automated classification, regression, and pattern recognition, while recent advances in physics-informed machine learning (PIML) have demonstrated that incorporating physical knowledge improves robustness and scientific consistency (Hao et al., 2022; Meng et al., 2025). Despite this progress, current models remain largely task-specific and depend heavily on labelled data, predefined physical constraints, or computationally expensive numerical simulations. The next generation of scientific AI must move beyond accurate prediction to develop systems capable of integrating observations, simulations, governing physics, uncertainty, and continuous learning within a unified framework.

7.1 Beyond Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) have established a strong foundation for combining machine learning with governing physical equations. However, their widespread adoption remains limited by high computational cost, optimisation instability, and reduced scalability for large multi-physics problems (Ghalambaz et al., 2024; Meng et al., 2025). Future research is therefore expected to focus on hybrid architectures that combine PINNs with neural operators, surrogate models, probabilistic learning, and symbolic reasoning. Such frameworks could retain physical consistency while improving computational efficiency and adaptability across different astronomical applications.

7.2 Scientific Foundation Models

Foundation models have transformed natural language processing and computer vision by learning transferable representations from large-scale datasets. A similar concept is emerging in scientific machine learning, where future models may be pretrained using heterogeneous astronomical data, including sky surveys, numerical simulations, synthetic observations, and physical constraints. Rather than being developed for individual tasks, these models could provide general scientific representations adaptable to stellar astrophysics, galaxy evolution, cosmology, gravitational-wave astronomy, and space-weather forecasting. Such transfer learning has the potential to reduce dependence on task-specific datasets while improving robustness across diverse scientific problems.

7.3 Multi-Fidelity Scientific Learning

Astronomical information is inherently heterogeneous, combining observations with different resolutions, numerical simulations of varying fidelity, and theoretical models with different levels of complexity. Future AI systems should exploit these complementary sources rather than treating them independently. Multi-fidelity learning provides a promising framework by integrating inexpensive simulations, high-resolution numerical models, observational data, and physical constraints within a single optimisation process. This strategy can improve predictive reliability while reducing computational cost and enabling efficient learning under limited observational data.

7.4 Digital Twins for Astronomy

Digital twins represent one of the most promising long-term directions for scientific AI by moving beyond static prediction towards continuously evolving computational models. Unlike conventional machine learning systems, digital twins assimilate new observations in real time to maintain an updated digital representation of a physical system. In astronomy, this concept could enable continuous monitoring of stellar evolution, solar activity, planetary systems, galaxies, and cosmological structures by integrating observational data, numerical simulations, uncertainty quantification, and machine learning within a unified computational framework (Sun et al., 2025; Zhou et al., 2026). Potential applications include adaptive telescope scheduling, mission planning, anomaly detection, predictive maintenance of space observatories, and long-term monitoring of dynamic astrophysical phenomena. By continuously comparing observations with physics-based simulations, digital twins could also improve parameter estimation, identify inconsistencies between models and observations, and support more efficient scientific decision-making. Despite their considerable potential, digital twins remain at an early stage of development in astronomy. Their implementation requires continuous access to high-quality observational data, reliable data assimilation techniques, scalable computational infrastructure, and robust uncertainty quantification. Integrating heterogeneous data from multiple telescopes and simulations while maintaining physical consistency also remains a significant challenge. Consequently, most current applications remain conceptual or limited to small-scale demonstration studies rather than operational astronomical systems. Rather than representing an immediate replacement for existing analysis workflows, digital twins should be viewed as a long-term evolution of scientific AI. Future progress will depend on advances in real-time data assimilation, high-performance computing, explainable AI, and physics-informed modelling. As these capabilities mature, digital twins have the potential to transform astronomy from retrospective data analysis towards continuously updated, predictive, and physically consistent scientific modelling.

7.5 AI-Assisted Scientific Discovery

Perhaps the most transformative direction is the development of AI systems capable of supporting scientific discovery itself. Rather than simply analysing existing observations, future systems may identify unexpected physical relationships, recommend new observations, optimise simulation parameters, detect anomalies, and generate testable scientific hypotheses. Importantly, this should not be viewed as replacing human researchers. Scientific progress depends on creativity, physical intuition, and critical judgement that remain uniquely human strengths. Instead, AI is likely to function as a collaborative research assistant, rapidly exploring large hypothesis spaces while allowing scientists to evaluate and interpret the resulting evidence.

7.6 Remaining Challenges

Several challenges must be addressed before trustworthy scientific AI becomes routine in astronomy. Computational scalability remains a major obstacle for many physics-informed models, particularly in large multi-scale simulations. Uncertainty quantification should become an integral component of model development rather than a post-processing step, enabling predictions to be accompanied by reliable confidence estimates.

Standardised benchmarking and independent validation are also required to enable objective comparison across competing methodologies. Finally, explainability must evolve beyond feature attribution towards physically meaningful interpretation, while open science practices—including shared datasets, reproducible workflows, and transparent reporting will be essential for building community confidence in future AI systems (Fluke & Jacobs, 2020; Hao et al., 2022). The evolution of artificial intelligence in astronomy can be viewed as a progression through four stages: **(i)** data-driven machine learning, **(ii)** physics-informed learning, **(iii)** integrated scientific machine learning, and **(iv)** autonomous scientific AI. Each stage represents a deeper integration of observational evidence, governing physics, numerical simulations, and computational intelligence. While fully autonomous scientific discovery remains a long-term objective, the most immediate opportunity lies in developing trustworthy AI systems that combine physical consistency, uncertainty awareness, explainability, and reproducibility within unified scientific workflows. Ultimately, the future of astronomical AI will depend not on replacing physics with machine learning but on integrating both to create computational systems that accelerate discovery while remaining scientifically reliable. Achieving this vision will require sustained collaboration between astronomers, physicists, computational scientists, and AI researchers, ensuring that artificial intelligence evolves from a powerful analytical tool into a trusted partner in scientific exploration. Building on the critical assessment presented throughout this review, Fig. 6 proposes a conceptual roadmap for the evolution of artificial intelligence in astronomy and astrophysics, illustrating the progressive transition from conventional data-driven machine learning to autonomous scientific AI through increasing integration of physical knowledge, scientific reasoning, and adaptive computational frameworks.

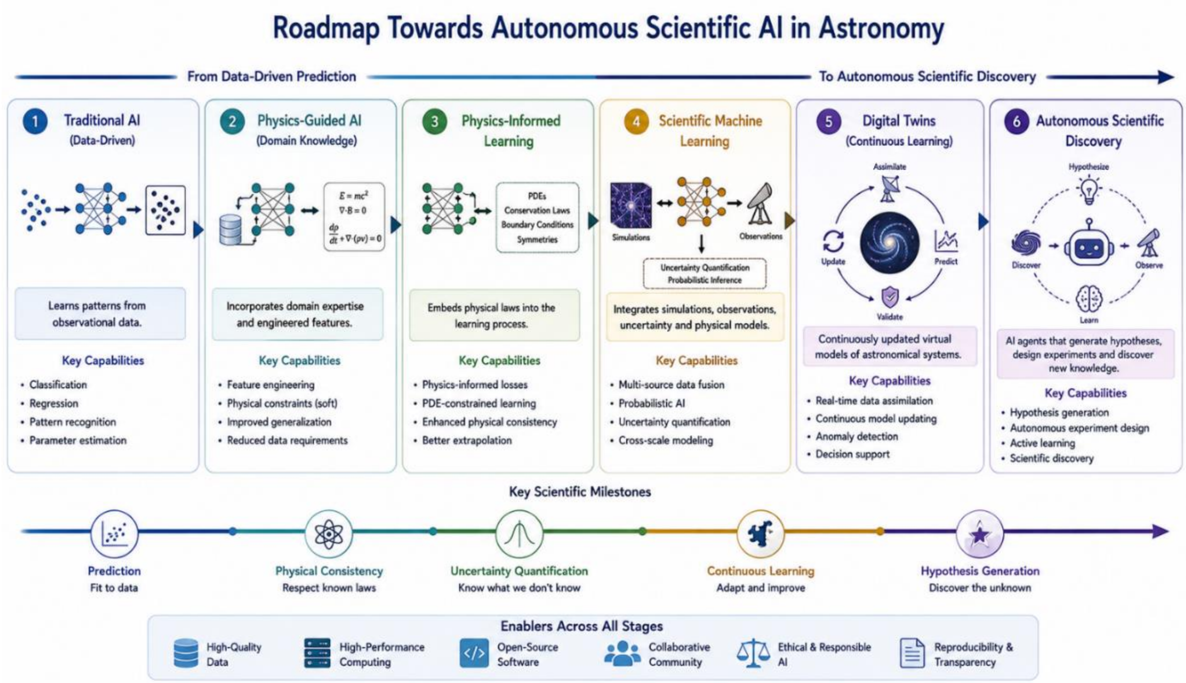

Fig. 6. Roadmap towards autonomous scientific artificial intelligence in astronomy and astrophysics

The figure illustrates the conceptual evolution of artificial intelligence from conventional data-driven machine learning to autonomous scientific AI. The progression begins with traditional machine learning, which primarily performs statistical prediction from observational data, and advances through physics-guided learning, physics-informed learning, and scientific machine learning, where increasing levels of physical knowledge, governing equations, numerical simulations, uncertainty quantification, and probabilistic inference are integrated into the learning process. The roadmap then extends to continuously learning digital twins that assimilate real-time observations and finally to autonomous scientific AI systems capable of supporting hypothesis generation, adaptive experiment design, active learning, and AI-assisted scientific discovery. Across all stages, progress depends on common enabling technologies, including high-quality observational data, high-performance computing, open-source software, collaborative scientific communities, ethical AI practices, and reproducible research workflows. The figure highlights the progressive transition from prediction-oriented AI towards physically consistent, trustworthy, and autonomous scientific intelligence for next-generation astronomical research. Figure developed by the author based on the synthesis of recent literature (Hao et al., 2022; Ghalambaz et al., 2024; Meng et al., 2025; Sun et al., 2025; Zhou et al., 2026). To summarize the future directions identified throughout this review, Table 6 presents a critical assessment of the major research priorities for advancing autonomous scientific AI in astronomy and astrophysics, highlighting their current maturity, scientific challenges, potential impact, and key areas requiring further investigation.

Table 6. Critical assessment of future research priorities for autonomous scientific AI in astronomy and astrophysics

Research Direction	Current Readiness	Major Scientific Challenge	Potential Scientific Impact	Key Research Priority
Physics-Informed Neural Networks (PINNs)	Mature	Poor scalability, optimisation instability, and computational cost for high-dimensional multi-physics systems	Physically consistent modelling of inverse problems, stellar interiors, plasma physics, and magnetohydrodynamics	Develop scalable optimisation strategies, adaptive training, and hybrid PINN architectures
Neural Operators	Emerging	Dependence on extensive high-fidelity simulation datasets and limited physical interpretability	Rapid surrogate modelling for cosmological simulations, radiative transfer, and computational astrophysics	Integrate operator learning with physical constraints and uncertainty quantification
Scientific Foundation Models	Early Emerging	Large-scale scientific pretraining, heterogeneous data integration, and domain adaptation	Transferable representations across multiple astronomical tasks with reduced task-specific training	Develop multimodal foundation models combining observations, simulations, and physical knowledge
Multi-Fidelity Learning	Emerging	Integrating observational data with simulations of different fidelity and uncertainty	Improved prediction accuracy, computational efficiency, and robustness under sparse observations	Establish unified frameworks for combining observational, simulation, and theoretical information
Digital Twins	Early Emerging	Continuous data assimilation, real-time model updating, and computational infrastructure	Adaptive modelling of dynamic astrophysical systems, mission planning, anomaly detection, and operational forecasting	Develop scalable digital twin architectures for observational astronomy
Autonomous Scientific AI	Conceptual	Reliable hypothesis generation, scientific reasoning, validation, and human–AI collaboration	AI-assisted discovery, adaptive experiment design, and accelerated scientific exploration	Combine symbolic reasoning, large foundation models, and physics-informed learning within trustworthy scientific workflows
Trustworthy Scientific AI	Developing	Explainability, uncertainty quantification, reproducibility, benchmarking, and transparent validation	Reliable deployment of AI for scientific inference and decision support	Establish community-wide standards for validation, benchmarking, uncertainty-aware learning, and open scientific practices

Critical Perspective. The future of astronomical AI will not be defined by a single breakthrough algorithm but by the convergence of complementary technologies. Physics-informed learning provides the foundation for physically consistent prediction, neural operators address computational scalability, foundation models improve transferability across scientific tasks, and digital twins enable continuous integration of observations with numerical simulations. Underpinning all these advances is the need for trustworthy AI through robust uncertainty quantification, explainability, reproducibility, and standardised benchmarking. Collectively, these research directions provide the technological pathway from task-specific machine learning towards autonomous scientific AI capable of supporting reliable astronomical discovery.

8. Conclusion

The convergence of modern astronomy and artificial intelligence is reshaping the way complex astrophysical phenomena are investigated. As astronomical observations and numerical simulations continue to increase in scale and complexity, the need for machine learning models that are not only accurate but also scientifically reliable has become increasingly important. Physics-informed machine learning addresses this challenge by embedding physical knowledge within data-driven algorithms, enabling models that better respect established scientific principles while improving robustness, interpretability, and generalisation. This review demonstrates that physics-informed learning should be viewed as a continuum of methodologies rather than a single modelling paradigm. By organising the literature according to the degree of physics integration, a clearer understanding emerges of how different approaches balance predictive capability, physical consistency, computational cost, and scientific applicability. This methodology-oriented perspective provides a common framework for comparing diverse techniques across astronomy and astrophysics and helps identify where specific approaches are most effective. The analysis also indicates that there is no universally optimal solution. The choice of methodology should be guided by the scientific question being addressed, the availability and quality of observational or simulated data, the complexity of the underlying physical processes, and the level of interpretability required. While conventional machine learning remains highly effective for many large-scale data analysis tasks, increasing levels of physics integration become particularly valuable in problems where physical constraints, limited observations, or reliable extrapolation are essential. At the same time, several challenges continue to limit broader adoption. The absence of standardised benchmark datasets, inconsistent validation practices, limited uncertainty quantification, and the high computational demands of many physics-informed approaches remain significant barriers to reproducible and comparable research. Addressing these issues will require greater methodological standardisation, transparent reporting, open benchmarking frameworks, and closer collaboration between astronomers, computational scientists, and artificial intelligence researchers. Emerging directions including neural operators, scientific foundation models, multi-fidelity learning, and digital twin technologies offer exciting opportunities for future research. However, their long-term scientific value will depend on rigorous validation against observational evidence and their ability to provide physically meaningful insights rather than incremental improvements in predictive accuracy alone. Ultimately, the future of artificial intelligence in astronomy is unlikely to be driven solely by increasingly sophisticated algorithms, but by the successful integration of data-driven learning with established physical understanding. By critically evaluating current methodologies, identifying their strengths and limitations, and proposing a unified framework for their assessment, this review provides a foundation for future research aimed at developing reliable, interpretable, and scientifically credible machine learning methods for next-generation astronomical discovery.

9. Limitations

Despite providing a comprehensive and methodology-oriented synthesis of recent advances in physics-informed machine learning (PIML) for astronomy and astrophysics, this review has several inherent limitations. First, the field is evolving at an exceptional pace, with new algorithms, benchmark datasets, observational resources, and computational frameworks being introduced continuously. As a result, some recently published developments may not be captured, despite the structured literature search undertaken for this review. This limitation is common to rapidly advancing interdisciplinary fields in which methodological innovation often outpaces the publication cycle.

Second, the objective of this review is to provide a critical methodological synthesis rather than a quantitative systematic review or meta-analysis. Accordingly, the emphasis is placed on evaluating conceptual advances, levels of physics integration, methodological strengths, scientific reliability, and emerging research directions instead of statistically comparing the predictive performance of individual models. Consequently, the conclusions should be interpreted as an evidence-based assessment of methodological trends rather than a definitive ranking of competing approaches.

Another important limitation is the considerable heterogeneity of the available literature. Existing studies differ substantially in terms of scientific objectives, observational datasets, numerical simulations, validation protocols, machine learning architectures, performance metrics, and reporting practices. Such variability limits direct quantitative comparison and makes it difficult to establish universally applicable conclusions regarding model performance or generalisation. Furthermore, several promising research directions including neural operators, scientific foundation models, digital twins, and adaptive scientific AI remain in the early stages of adoption within astronomy and astrophysics. Current evidence is therefore derived primarily from proof-of-concept investigations or domain-specific applications, and their broader scientific value will require validation across a wider range of astrophysical problems. Finally, the review highlights a broader challenge facing the community: the absence of standardised benchmarking datasets, harmonised evaluation protocols, and widely accepted reporting guidelines for physics-informed machine learning in astronomy. Greater methodological standardisation, transparent reporting, reproducible computational workflows, and systematic cross-domain validation will be essential for enabling rigorous comparison between competing approaches and accelerating the translation of PIML from promising research methodology to a reliable scientific tool. Within this context, the present review should be viewed as a critical snapshot of the current state of the field and a framework for guiding future research, rather than as a final or exhaustive assessment of this rapidly evolving discipline.

Final Perspective

Astronomy is entering a new phase in which artificial intelligence is evolving from a powerful data-analysis tool into a collaborative scientific framework. Future advances will depend on developing AI systems that integrate observations, simulations, governing physics, and uncertainty-aware reasoning within transparent and reproducible workflows. Rather than replacing scientific expertise, these systems have the potential to augment human reasoning, accelerate discovery, and provide deeper insight into the physical processes that govern the Universe. Achieving this vision will define the next generation of trustworthy scientific AI for astronomy and astrophysics.

Acknowledgements

The author gratefully acknowledges the researchers whose published work has contributed to the development of this review. The author also thanks the scientific community for advancing research in artificial intelligence and astronomy, which provided the foundation for this critical synthesis.

Declaration of AI Use

This manuscript was prepared through the original scholarly contributions of the author(s), including the study design, literature analysis, critical interpretation, content development, and manuscript preparation. Grammarly was used solely to assist with English language editing, grammar correction, and stylistic refinement. No artificial intelligence tool was used to generate or interpret the scientific content, formulate conclusions, or make scholarly decisions presented in this manuscript. The author(s) carefully reviewed and approved all manuscript content and accept full responsibility for its accuracy, integrity, and scientific validity.

All figures and graphical illustrations presented in this manuscript are original conceptual illustrations developed by the author based on a critical synthesis of the reviewed literature and the author's scientific interpretation and perspective. Vector illustration and graphic design software were used solely to prepare the visual presentation of these concepts. The figures do not reproduce copyrighted material from previously published sources and are intended to communicate the author's original conceptual framework developed from the literature reviewed in this article.

Competing Interests

Author has declared that no known competing financial interests or non-financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbar, S., Wu, M.-R., & Xiong, Z. (2024). Physics-informed neural networks for predicting the asymptotic outcome of fast neutrino flavor conversions. *Physical Review D, 109*(4), Article 043024. https://doi.org/10.1103/PhysRevD.109.043024

Audenaert, J. (2025). From stellar light to astrophysical insight: Automating variable star research with machine learning. *Astrophysics and Space Science, 370*, Article 72. https://doi.org/10.1007/s10509-025-04460-5

Ball, N. M., & Brunner, R. J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D, 19*(07), 1049–1106. https://doi.org/10.1142/S0218271810017160

Baron, D. (2019). Machine learning in astronomy: A practical overview. *arXiv*. https://doi.org/10.48550/arXiv.1904.07248

Baty, H. (2023). Modelling Lane–Emden type equations using physics-informed neural networks. *Astronomy and Computing, 44*, Article 100734. https://doi.org/10.1016/j.ascom.2023.100734

Baty, H. (2024). A hands-on introduction to physics-informed neural networks for solving partial differential equations with benchmark tests taken from astrophysics and plasma physics. *arXiv*. https://doi.org/10.48550/arXiv.2403.00599

Bellinger, E. P., Angelou, G. C., Hekker, S., Basu, S., Ball, W. H., & Guggenberger, E. (2016). Fundamental parameters of main-sequence stars in an instant with machine learning. *The Astrophysical Journal, 830*(1), Article 31. https://doi.org/10.3847/0004-637X/830/1/31

Buchner, J., & Fotopoulou, S. (2024). How to set up your first machine learning project in astronomy. *Nature Reviews Physics, 6*, 535–545. https://doi.org/10.1038/s42254-024-00743-y

Dai, Z., Moews, B., Vilalta, R., & Davé, R. (2024). Physics-informed neural networks in the recreation of hydrodynamic simulations from dark matter. *Monthly Notices of the Royal Astronomical Society, 527*(2), 3381–3394. https://doi.org/10.1093/mnras/stad3394

Fluke, C. J., & Jacobs, C. (2020). Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(2), Article e1349. https://doi.org/10.1002/widm.1349

Fotopoulou, S. (2024). A review of unsupervised learning in astronomy. *Astronomy and Computing, 48*, Article 100851. https://doi.org/10.1016/j.ascom.2024.100851

Ghalambaz, M., Sheremet, M. A., Khan, M. A., Raizah, Z., & Shafi, J. (2024). Physics-informed neural networks (PINNs): Application categories, trends and impact. *International Journal of Numerical Methods for Heat & Fluid Flow, 34*(8), 3131–3165. https://doi.org/10.1108/HFF-09-2023-0568

Hao, Z., Liu, S., Zhang, Y., Ying, C., Feng, Y., Su, H., & Zhu, J. (2022). Physics-informed machine learning: A survey on problems, methods and applications. *arXiv*. https://doi.org/10.48550/arXiv.2211.08064

Jarolim, R., Thalmann, J. K., Veronig, A. M., & Podladchikova, T. (2023). Probing the solar coronal magnetic field with physics-informed neural networks. *Nature Astronomy, 7*(10), 1171–1179. https://doi.org/10.1038/s41550-023-02030-9

Kugel, R., Schaye, J., Schaller, M., Helly, J. C., Braspenning, J., Elbers, W., Frenk, C. S., McCarthy, I. G., Kwan, J., Salcido, J., van Daalen, M. P., Vandenbroucke, B., Bahé, Y. M., Borrow, J., Chaikin, E., Huško, F., Jenkins, A., Lacey, C. G., Nobels, F. S. J., ... Vernon, I. (2023). FLAMINGO: Calibrating large cosmological hydrodynamical simulations with machine learning. *Monthly Notices of the Royal Astronomical Society, 526*(4), 6103–6127. https://doi.org/10.1093/mnras/stad2540

Kumaran, S., Mandal, S., & Bhattacharyya, S. (2025). Explainable machine learning classification of Chandra X-ray sources: SHAP analysis of multiwavelength features. *The Astrophysical Journal, 995*(2), Article 224. https://doi.org/10.3847/1538-4357/ae1a45

Li, G., Lu, Z., Wang, J., & Wang, Z. (2025). Machine learning in stellar astronomy: Progress up to 2024. *arXiv*. https://doi.org/10.48550/arXiv.2502.15300

Li, M., Liu, S., Wang, J., Cui, Y., Luo, B., & Wang, X. (2026). A physics-informed deep learning model for the magnetic classification of solar active regions. *The Astrophysical Journal Supplement Series, 283*(2), Article 44. https://doi.org/10.3847/1538-4365/ae4595

Lu, X., & Wang, Y. (2024). Surrogate modeling for radiative heat transfer using physics-informed deep neural operator networks. *Proceedings of the Combustion Institute, 40*(1–4), Article 105282. https://doi.org/10.1016/j.proci.2024.105282

Marín Díaz, G. (2025). Integrating exploratory data analysis and explainable AI into astronomy education: A fuzzy approach to data-literate learning. *Education Sciences, 15*(12), Article 1688. https://doi.org/10.3390/educsci15121688

Mbouopda, M. F., Ishida, E. E. O., Nguifo, E. M., & Gangler, E. (2026). Explainable classification of astronomical uncertain time series. *Big Data Research, 43*, Article 100591. https://doi.org/10.1016/j.bdr.2026.100591

Meher, S. K., & Panda, G. (2021). Deep learning in astronomy: A tutorial perspective. *The European Physical Journal Special Topics, 230*(10), 2285–2317. https://doi.org/10.1140/epjs/s11734-021-00207-9

Meng, C., Griesemer, S., Cao, D., Seo, S., & Liu, Y. (2025). When physics meets machine learning: A survey of physics-informed machine learning. *Machine Learning for Computational Science and Engineering, 1*, Article 20. https://doi.org/10.1007/s44379-025-00016-0

Meray, A., Wang, L., Kurihana, T., Mastilovic, I., Praveen, S., Xu, Z., Memarzadeh, M., Lavin, A., & Wainwright, H. (2024). Physics-informed surrogate modeling for supporting climate resilience at groundwater contamination sites. *Computers & Geosciences, 183*, Article 105508. https://doi.org/10.1016/j.cageo.2023.105508

Mohammadi, M., Mutatiina, J., Saifollahi, T., & Bunte, K. (2022). Detection of extragalactic ultra-compact dwarfs and globular clusters using explainable AI techniques. *Astronomy and Computing, 39*, Article 100555. https://doi.org/10.1016/j.ascom.2022.100555

Mombarg, J. S. G., Van Reeth, T., & Aerts, C. (2021). Constraining stellar evolution theory with asteroseismology of γ Doradus stars using deep learning: Stellar masses, ages, and core-boundary mixing. *Astronomy & Astrophysics, 650*, Article A58. https://doi.org/10.1051/0004-6361/202140835

Moschou, S. P., Hicks, E., Parekh, R. Y., Mathew, D., Majumdar, S., & Vlahakis, N. (2023). Physics-informed neural networks for modeling astrophysical shocks. *Machine Learning: Science and Technology, 4*(3), Article 035032. https://doi.org/10.1088/2632-2153/acf116

Okwuwe, J., & Oduselu-Hassan, O. E. (2024). AI-augmented finite difference methods for solving PDEs: Advancing numerical solutions in mathematical modeling. *Asian Journal of Mathematics and Computer Research, 31*(4), 56–67. https://doi.org/10.56557/AJOMCOR/2024/v31i48985

Pilania, G., McClellan, K. J., Stanek, C. R., & Uberuaga, B. P. (2018). Physics-informed machine learning for inorganic scintillator discovery. *The Journal of Chemical Physics, 148*(24), Article 241729. https://doi.org/10.1063/1.5025819

Sen, S., Agarwal, S., Chakraborty, P., & Singh, K. P. (2022). Astronomical big data processing using machine learning: A comprehensive review. *Experimental Astronomy, 53*(1), 1–43. https://doi.org/10.1007/s10686-021-09827-4

Singh, P., Del Rose, T., Palasyuk, A., & Mudryk, Y. (2023). Physics-informed machine-learning prediction of Curie temperatures and its promise for guiding the discovery of functional magnetic materials. *Chemistry of Materials, 35*(16), 6304–6312. https://doi.org/10.1021/acs.chemmater.3c00892

Sun, L., Hu, Z., Todd, M. D., & Zeng, J. (2025). Generative artificial intelligence for Bayesian model updating in digital twins: A review and tutorial. *Structural and Multidisciplinary Optimization, 68*, Article 249. https://doi.org/10.1007/s00158-025-04183-9

Ting, Y. S. (2025). Deep learning in astrophysics. arXiv. https://doi.org/10.48550/arXiv.2510.10713

Urbán, J. F., Stefanou, P., Dehman, C., & Pons, J. A. (2023). Modelling force-free neutron star magnetospheres using physics-informed neural networks. *Monthly Notices of the Royal Astronomical Society, 524*(1), 32–42. https://doi.org/10.1093/mnras/stad1810

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Spergel, D. N., Somerville, R. S., Davé, R., Pillepich, A., Hernquist, L., Nelson, D., Torrey, P., Narayanan, D., Li, Y., Philcox, O., La Torre, V., Delgado, A. M., Ho, S., Hassan, S., Burkhart, B., Wadekar, D., ... Bryan, G. L. (2021). The CAMELS project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal, 915*(1), Article 71. https://doi.org/10.3847/1538-4357/abf7ba

Wayo, D. D. K. (2026). Ensembles of graph and physics-informed machine learning for scientific modeling in materials science: A review. *Archives of Computational Methods in Engineering, 33*(1), 963–988. https://doi.org/10.1007/s11831-025-10325-5

Zhai, W. (2026). Bridge moving load identification: A state-of-the-art review. *Advances in Research, 27*(1), 251–263. https://doi.org/10.9734/AIR/2026/v27i11585

Zhou, R., Chen, D., Jia, Z., Su, Y., Liu, Y., Lu, Y., Shi, D., Huang, Y., Xu, T., Pan, Y., Li, X., Abate, Y., Chen, Q., Tu, Z., Yang, Y., Zhang, Y., Wen, Q., Mai, G., Fu, S., ... He, L. (2026). Digital twin AI: Opportunities and challenges from large language models to world models. *arXiv*. https://doi.org/10.48550/arXiv.2601.01321

Appendix

To provide an integrated overview of the concepts discussed throughout this review, Fig. A1 presents a graphical synthesis of the methodological evolution, key enabling technologies, major application areas, and future research directions that underpin the development of trustworthy scientific artificial intelligence in astronomy and astrophysics.

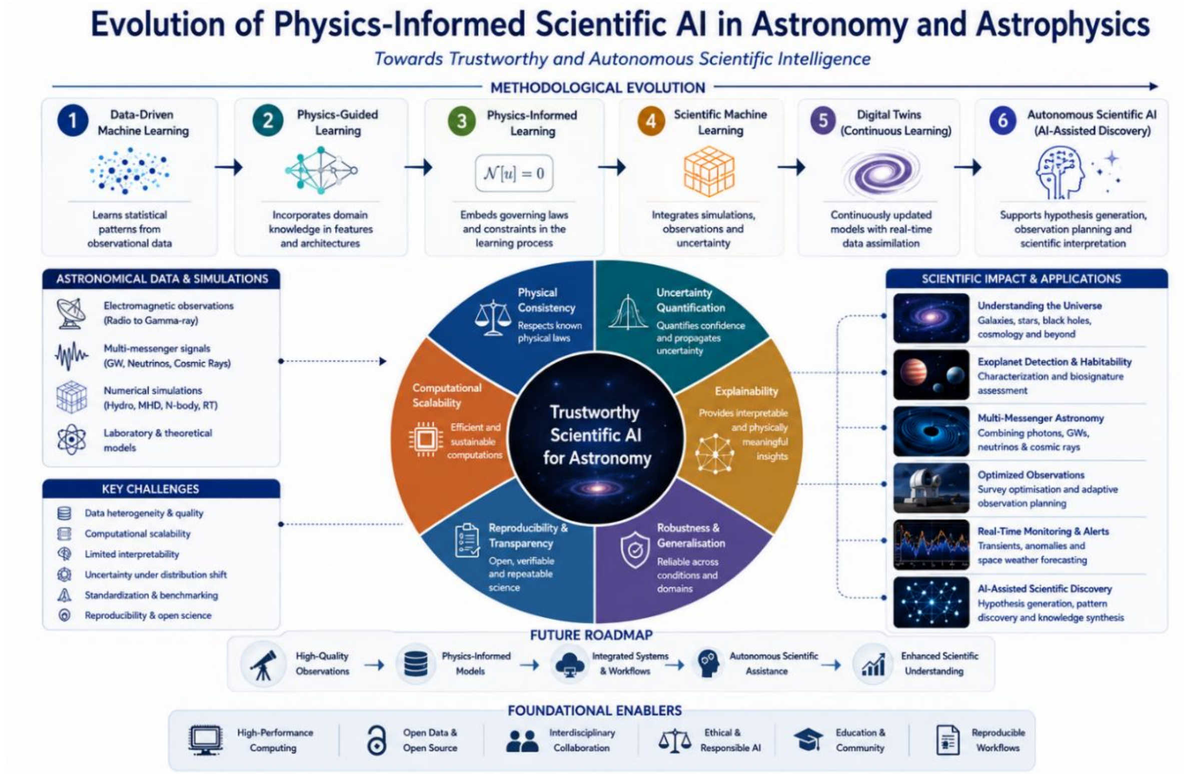

Fig. A1. Graphical summary of the evolution towards trustworthy scientific artificial intelligence in astronomy and astrophysics

The figure synthesizes the principal concepts discussed throughout this review, illustrating the progression from conventional data-driven machine learning to physics-guided learning, physics-informed learning, scientific machine learning, digital twins, and autonomous scientific AI. It highlights the integration of heterogeneous astronomical observations, numerical simulations, governing physical laws, uncertainty quantification, explainability, and continuous learning within unified computational frameworks. The central framework emphasises that trustworthy scientific AI depends on six interconnected principles: physical consistency, uncertainty quantification, explainability, robustness and generalisation, reproducibility and transparency, and computational efficiency. The figure further summarises the major scientific applications, current methodological challenges, enabling technologies, and future research priorities that will shape the next generation of AI-assisted astronomical discovery. Author-developed graphical summary based on the synthesis of the reviewed literature.

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/162019