International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Artificial Intelligence for Scientific Discovery: A Review Of Methods, Landmark Applications and The Emergence Of Autonomous Research

Lende Vedika Vijay[1] Prof. R. V. Varhade[2] Prof. M. C. Kolse[3] Prof. S. K. Said4
Department of Artificial Intelligence and Data Science
Jaihind College of Engineering
Pune, Maharashtra, India

Abstract—Artificial Intelligence is transforming how science is done by helping researchers analyse complex data, uncover hidden patterns, generate hypotheses and accelerate discovery across disciplines. This review examines the role of AI in scientific discovery, covering methods, applications, benefits, challenges and future possibilities. Machine learning, deep learning, physics-informed neural networks, equation discovery, natural language processing of scientific literature, reinforcement learning and generative AI are explained in terms of how they support data-driven investigation. Landmark applications are reviewed across structural biology, drug and antibiotic discovery, chemistry, materials science, physics, fusion-energy control, mathematics, astronomy and weather and climate forecasting. The emergence of autonomous laboratories and language-model agents that plan and execute experiments is discussed. A five-layer framework for AI-assisted discovery comprising data, representation, modelling, hypothesis generation and experimental validation is proposed, with research gaps, objectives, a comparison of AI roles and challenges including data quality, interpretability, reproducibility, physical consistency and the risk of over-reliance. The review concludes that AI is most powerful as a partner to scientists, accelerating the cycle of hypothesis, experiment and understanding while human judgement guides what is meaningful.

Index Terms—AI for science, scientific discovery, physics-informed neural networks, protein structure prediction, materials discovery, autonomous laboratories

I. INTRODUCTION

Scientific progress has always depended on observation, hypothesis and experiment. Today, instruments generate data far faster than humans can examine it: telescopes survey billions of objects, sequencers read entire genomes and simulations produce petabytes of results. At the same time, the space of possible molecules, materials and designs is too large to explore by hand.

AI helps in both directions: it extracts patterns from massive data and proposes promising candidates for experiments. Wang *et al.* [1] reviewed scientific discovery in the age of AI and described how learning methods assist with data collection, hypothesis generation, experimentation and interpretation.

This paper reviews AI methods and landmark applications in science and proposes a framework for AI-assisted discovery.

II. BACKGROUND AND RELATED WORK

2.1 Roles Of Ai In Science

Fig. 1 summarizes the roles: analysing data, predicting properties, discovering equations, generating hypotheses and candidates, and automating experiments.

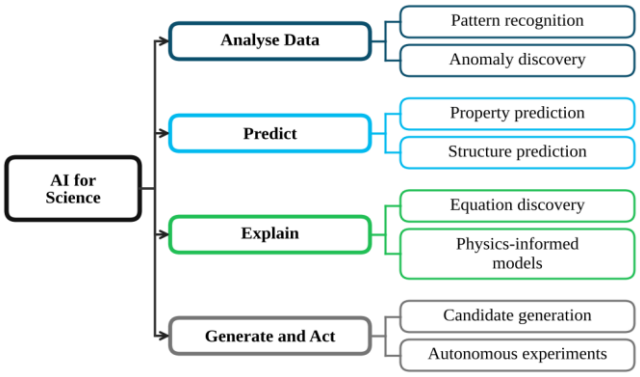


Fig. 1. Roles of AI in the scientific process.

2.2 Physics-Informed Learning

Raissi *et al.* [2] introduced physics-informed neural networks that embed differential equations in the loss function, so models respect known physical laws while learning from data. Brunton *et al.* [3] discovered governing equations directly from data using sparse regression.

2.3 Review Of Related Studies

Biology and medicine. AlphaFold [4] predicted protein structures with near-experimental accuracy, and Stokes *et al.* [5] used a graph neural network to discover a new antibiotic.

© 2026 IJSRET

Materials and chemistry. Merchant *et al.* [6] used graph networks to predict stable crystal structures at very large scale, and Szymanski *et al.* [7] demonstrated an autonomous laboratory that synthesized new inorganic materials with minimal human intervention. Tshitoyan *et al.* [8] showed that word embeddings trained on materials-science abstracts captured latent knowledge and could suggest materials for applications before they were reported. Boiko *et al.* [9] built a language-model agent that planned and executed chemistry experiments.

Physics and engineering. Baldi *et al.* [10] showed deep networks improve the search for exotic particles in collider data, and Degrave *et al.* [11] used reinforcement learning to control plasma shape in a tokamak fusion reactor.

Mathematics and computation. Fawzi *et al.* [12] discovered faster matrix multiplication algorithms with reinforcement learning, and Romera-Paredes *et al.* [13] combined language models with evaluators in FunSearch to discover new mathematical constructions.

Earth science. GraphCast [14] produced skilful medium-range weather forecasts faster than traditional numerical models. Table I summarizes these works.

Ref.	Authors	Year	Technique Used	Merits	Demerits
[8]	Tshitoyan et al.	2019	Word embeddings of literature	Predicted materials before discovery	Text-only knowledge
[2]	Raissi et al.	2019	Physics-informed neural networks	Embed physical laws in learning	Training difficulties
[4]	Jumper et al.	2021	AlphaFold structure prediction	Near-experimental protein structures	Static structures
[11]	Degrave et al.	2022	Deep RL tokamak control	Controlled plasma configurations	Single device
[12]	Fawzi et al.	2022	RL algorithm discovery	Faster matrix multiplication	Narrow problem
[6]	Merchant et al.	2023	Graph networks for crystals (GNoME)	Large-scale stable material prediction	Synthesis not guaranteed
[7]	Szymanski et al.	2023	Autonomous materials laboratory	Synthesized new materials autonomously	Specific chemistry
[13]	Romera-Paredes et al.	2024	LLM program search (FunSearch)	New mathematical constructions	Needs evaluator

TABLE I Summary of Literature Survey

III. RESEARCH GAP AND OBJECTIVES

3.1 Identified Research Gaps

• Many AI predictions are not experimentally validated, limiting scientific confidence.

• Black-box models produce answers without mechanistic understanding.

• Scientific data are often scarce, noisy or biased toward well-studied systems.

• Reproducibility is hindered by unreleased code, data and compute requirements.

• Access to AI for science is unequal across institutions and countries.

3.2 Objectives Of The Review

- **O1:** To review AI methods used in scientific discovery.
- **O2:** To examine landmark applications across disciplines.
- **O3:** To propose a framework for AI-assisted discovery with experimental validation.
- **O4:** To compare AI roles in the scientific process.
- **O5:** To discuss challenges and future possibilities, including autonomous research.

IV. DESIGN OF THE PROPOSED SYSTEM

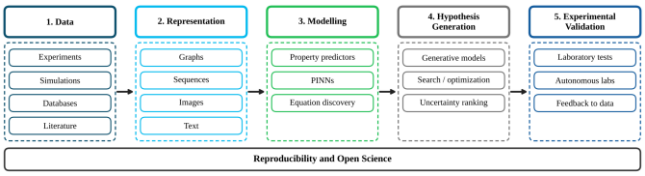


Fig. 2 presents the proposed framework.

Fig. 2. Proposed framework for AI-assisted scientific discovery.

4.1 Data Layer

Experimental measurements, simulations, databases and literature are collected and curated with metadata.

© 2026 IJSRET

4.2 Representation Layer

Data are encoded as graphs (molecules, crystals), sequences (proteins, genomes), images (microscopy, telescopes) or text (papers).

4.3 Modelling Layer

Predictive models estimate properties; physics-informed and equation-discovery models capture underlying laws.

4.4 Hypothesis Generation Layer

Generative models and search algorithms propose candidates and hypotheses ranked by predicted value and uncertainty.

4.5 Experimental Validation Layer

Laboratories, possibly automated, test top candidates; results feed back into the data layer, closing the discovery loop.

4.6 Methodology

Fig. 3 describes the discovery loop. A physics-informed neural network minimizes a combined loss

$$L = L_data + \lambda \cdot L_physics \quad (1)$$

where L_physics penalizes violations of the governing equations. Candidate selection often balances predicted value μ and uncertainty σ through an acquisition function such as the upper confidence bound

$$UCB(x) = \mu(x) + \kappa \cdot \sigma(x) \quad (2)$$

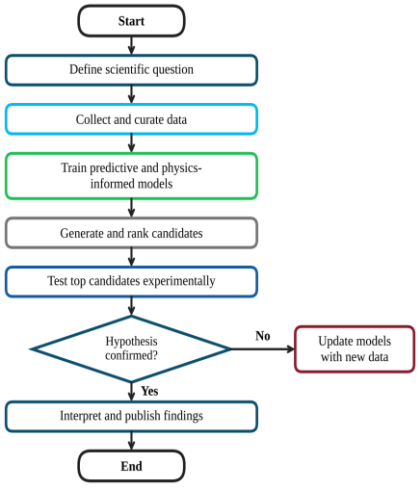


Fig. 3. AI-assisted discovery loop.

V. DISCUSSION AND COMPARISON

Table II compares AI roles in science. Data analysis and property prediction are mature, equation discovery adds interpretability, generative design proposes novel candidates and autonomous labs accelerate experimentation.

AI Role	Maturity	Interpretability	Example
Data analysis	High	Medium	Particle physics, astronomy
Property prediction	High	Low–Medium	Protein / material properties
Equation discovery	Medium	High	SINDy, PINNs
Generative design	Medium	Low	Molecules, crystals
Autonomous labs	Emerging	Medium	Robotic synthesis

TABLE II Comparison of AI Roles in Science

VI. CHALLENGES AND FUTURE DIRECTIONS

Table III summarizes challenges. Future directions include scientific foundation models, AI agents that read literature and design experiments, open shared datasets and compute for researchers in developing countries, and standards for reporting AI-assisted findings.

Challenge	Mitigation
Lack of validation	Closed loop with experiments
Black-box models	Physics-informed, interpretable methods
Scarce, noisy data	Transfer learning; data sharing
Reproducibility	Open code, data and benchmarks
Unequal access	Shared compute and open models

TABLE III Challenges and Mitigation

VII. CONCLUSIONS

AI has become a powerful partner in scientific research, delivering breakthroughs in protein structure, materials, antibiotics, fusion control, mathematics and weather forecasting. The reviewed literature shows that the greatest impact comes when AI is embedded in a loop with experiments and physical knowledge. Ensuring validation, interpretability and reproducibility, while keeping scientists in charge of what questions matter, will determine how far AI can accelerate discovery.

© 2026 IJSRET

7.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

REFERENCES

1. H. Wang et al., "Scientific discovery in the age of artificial intelligence," Nature, vol. 620, pp. 47–60, 2023.
2. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys., vol. 378, pp. 686–707, 2019.
3. S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl. Acad. Sci. USA, vol. 113, no. 15, pp. 3932–3937, 2016.
4. J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," Nature, vol. 596, no. 7873, pp. 583–589, 2021.
5. J. M. Stokes et al., "A deep learning approach to antibiotic discovery," Cell, vol. 180, no. 4, pp. 688–702, 2020.
6. A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," Nature, vol. 624, pp. 80–85, 2023.
7. N. J. Szymanski et al., "An autonomous laboratory for the accelerated synthesis of novel materials," Nature, vol. 624, pp. 86–91, 2023.
8. V. Tshitoyan et al., "Unsupervised word embeddings capture latent knowledge from materials science literature," Nature, vol. 571, pp. 95–98, 2019.
9. D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes, "Autonomous chemical research with large language models," Nature, vol. 624, pp. 570–578, 2023.
10. P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," Nat. Commun., vol. 5, Art. no. 4308, 2014.
11. J. Degrave et al., "Magnetic control of tokamak plasmas through deep reinforcement learning," Nature, vol. 602, pp. 414–419, 2022.
12. A. Fawzi et al., "Discovering faster matrix multiplication algorithms with reinforcement learning," Nature, vol. 610, pp. 47–53, 2022.
13. B. Romera-Paredes et al., "Mathematical discoveries from program search with large language models," Nature, vol. 625, pp. 468–475, 2024.
14. R. Lam et al., "Learning skillful medium-range global weather forecasting," Science, vol. 382, no. 6677, pp. 1416–1421, 2023.

© 2026 IJSRET