Journal of Data Science 24 (1), 106–124
January 2026
DOI: 10.6339/25-JDS1214
Data Science Reviews

AI for Science: Opportunities, Challenges, and Future Directions

Valerie Fu[1],*
[1]*Independent Researcher, Carmel, IN 46032, U.S.A.*

Abstract

Artificial intelligence (AI) has lately emerged as a transformative force in scientific discovery, with skills in accelerating knowledge synthesis, automating experimentation, and enhancing interdisciplinary collaboration. As research challenges—ranging from climate change to rare disease treatments—grow more and more complex, the rapid evolution of AI calls for a comprehensive examination of its current and future roles. Despite recent breakthroughs, the field remains fragmented, due to the lack of a unified framework to understand AI's progression in science and its implications for data science, in particular. To address this gap, this review provides an analysis on AI for science, and also introduces a novel three-phase framework—Keplerian (data-driven pattern recognition), Edisonian (autonomous experimentation), and Einsteinian (foundational innovation)—to conceptualize AI's evolving role in science. Additionally, we discuss the ethical, environmental, and data privacy challenges that go alongside AI's integration in science, emphasizing the need for sustainable and responsible development. This review outlines how AI may transform the scientific methods and to help researchers harness AI's potential to drive scientific innovation.

Keywords *Einsteinian phase AGI; human computer collaboration; knowledge dissemination; machine learning; scientific discovery; transdisciplinary research*

1 Introduction

Scientific research stands at a pivotal moment in history, characterized by both unprecedented potential and daunting challenges. Over the past half-century, the number of scientists in the US has increased sevenfold, yet the rate of transformative discoveries has not nearly kept up proportionally (Burke, 2019). The complexities within modern science have reached a point where progress is often hindered by bottlenecks in knowledge synthesis, resource limitations, and the sheer scale of interdisciplinary expertise required to address these challenges (MacLeod et al., 2019). Problems like combating climate change, decoding the human genome, developing treatments for rare diseases, and understanding socioeconomic systems demand deep specialization across multiple fields, which often takes decades to achieve (National Science Board, 2020; Leontidis, 2024; Griffin et al., 2024). This divergence between growth in scientific capability and the rate of impactful discoveries reflects the increasing difficulty of navigating vast amounts of information. The modern scientific landscape requires researchers to work through an overwhelming amount of literature, manage data from various sources, and collaborate across multiple fields to develop innovative solutions. As these fields become more specialized, divides

---

* Email: valerie.r.fu@gmail.com.

Received August 15, 2025; Accepted December 12, 2025

intensify, creating barriers to addressing complex, transdisciplinary problems (World Economic Forum, 2021). This paradox raises the question: *how can we accelerate the rate of discovery and overcome these systemic challenges?*

Artificial intelligence (AI) has since emerged as an importance force is transforming the scientific process (Kellogg Insight, 2024). By synthesizing knowledge across scientific fields and enabling simulations of complex systems, AI presents newfound opportunities in overcoming existing bottlenecks in research, discovery, and innovation (Google, 2024). Recent breakthroughs demonstrate this potential, highlighting a growing trend towards leveraging technology and AI to introduce solutions aimed at solving pressing scientific challenges (Griffin et al., 2024). Beyond automation, AI is also reshaping how knowledge is disseminated and applied. Traditional scientific publications, while essential for peer review and archival purposes, are often inaccessible to policymakers, educators, and the greater population (Youvan, 2024). AI-driven tools can bridge this gap by translating complex research findings into actionable steps for diverse audiences, thus fostering collaboration while accelerating the uptake of scientific knowledge (Reddy and Mathur, 2025). For instance, AI-powered platforms like Semantic Scholar use natural language processing to automatically generate summaries of scholarly papers, assisting researchers and policymakers in staying current with developments across various fields. Furthermore, advancements in machine translation, powered by AI, have made scientific literature more accessible to non-specialists by translating complex research findings into multiple languages, thus broadening the reach and impact of scientific discoveries (Steigerwald et al., 2022). Together, these AI-driven innovations are reshaping how knowledge is disseminated and applied, ensuring that scientific insights are more accessible and actionable for a wider audience.

However, the integration of AI into science is not without its challenges; concerns over data biases (Mehrabi et al., 2021), the environmental impact of computationally intensive models (Strubell et al., 2020), and ethical considerations surrounding dual-use applications (Brundage et al., 2018) emphasize the need for responsible AI development. Addressing these challenges requires an approach which combines technical innovation with ethical oversight and public engagement (Díaz-Rodríguez et al., 2023; Drew, 2024).

This review paper examines the role of AI in science, focusing on its current contributions, future opportunities, and associated risks. Section 2 discusses the impact of AI on modern science, highlighting breakthrough applications and their significance. Additionally, this section examines prospective opportunities, including AI's potential to accelerate discovery and address complex challenges in science. Special attention is given to AI's influence on data science, reviewing both its current impact and future possibilities. Section 3 addresses the limitations and risks of relying on AI, including its environmental impact and ethical concerns. Section 4 synthesizes insights and outlines potential future directions. Finally, Section 5 presents an overview and call to action. By incorporating recent advancements and landmark case studies, this review aims to provide a comprehensive and forward-looking perspective on AI's impact on the scientific landscape.

2 AI in Modern Science and Its Transformative Potential

In recent years, the growing complexity of scientific challenges has demanded interdisciplinary approaches and intensive computational resources. Within this context, AI has proven to be a key tool by breaking traditional barriers and accelerating advancements across various scientific fields. This section highlights specific instances that demonstrate AI's promising role in scientific discovery. Subsequently, we discuss the potential AI holds for scientific research, with emphasis on how it can catalyze breakthroughs across different disciplines.

2.1 Breakthrough Applications

The integration of AI into various scientific fields has revolutionized protein science, exemplified by DeepMind's AlphaFold 2 (Jumper et al., 2021) and AlphaFold 3 (Abramson et al., 2024). These programs have solved the decades-old protein folding problem by accurately predicting three-dimensional protein structures using a deep learning framework. Contrary to traditional approaches that rely heavily on chemistry and physics principles, AlphaFold employs a data-driven approach; it operates end-to-end, processing amino acid sequences as input and directly outputting 3D structures through a sophisticated neural network. This paradigm shift has not only advanced our understanding of protein structures but also catalyzed developments in related fields, enabling the identification of novel therapeutic targets and accelerating drug discovery processes. In a related advancement, biochemist and computational biologist David Baker's introduction of RFDiffusion—a method that combines structure prediction with generative diffusion models—has further expanded the potential of protein design (Watson et al., 2023). RFDiffusion facilitates the design of novel proteins that could revolutionize fields such as medicine and materials science by enabling the development of new therapies, vaccines, and advanced biomaterials. The collective contributions of DeepMind's team, including John Jumper and Demis Hassabis, alongside David Baker's pioneering work, were recognized with the 2024 Nobel Prize in Chemistry. This distinction underscores the vital role of computational methodologies in deciphering complex biological challenges, showcasing the emergence of what NVidia CEO Jensen Huang termed at GTC 2024 as "Digital Biology". According to Huang (2024), "For the first time in human history, biology has the opportunity to be engineered, not just studied as a science. When something shifts to an engineering discipline, it becomes more systematic and experiences exponential improvement. This allows for the compounding of benefits from previous years, and each researcher's contribution builds upon those of others in a significant way for the first time." This transition signifies a shift from traditional study to a more engineered approach, promising systematic improvements and accelerating scientific achievements. This transformation not only fosters cumulative advancement in biology but also enhances the scope of each scientist's contribution to an ever-evolving body of knowledge.

Beyond biology, this trend of AI-driven discovery extends to other fundamental areas of science. DeepMind's AlphaTensor project demonstrates the power of AI to discover novel and more efficient algorithms. AlphaTensor addressed the fundamental problem of matrix multiplication, an operation in computing with widespread applications in fields like machine learning and computer graphics. By framing the search for efficient matrix multiplication algorithms as a game, AlphaTensor, using reinforcement learning, discovered algorithms that outperform existing human-designed approaches for specific matrix sizes. This achievement goes beyond just improving computational efficiency: it showcases the potential of AI to uncover previously unknown mathematical structures and optimize core computational processes, impacting a wide range of scientific and technological endeavors (Fawzi et al., 2022). DeepMind's GraphCast model highlights how AI can go as far as revolutionizing weather forecasting. Traditional numerical weather prediction models are computationally intensive and can be slow, limiting their ability to provide timely and accurate forecasts, especially for extreme weather events. GraphCast employs a graph neural network to learn from historical weather data and make accurate short-to-medium range weather forecasts significantly faster than traditional methods. This speed advantage allows for more frequent updates and improved predictions of extreme weather, potentially saving lives and reducing the damage of natural disasters. This application of AI to a complex, real-world problem like weather forecasting demonstrates its abilities to address societal challenges and enhance our understanding of complex systems (Lam et al., 2023). Examples like AlphaTensor in mathematics, GraphCast in meteorology, and AlphaFold and RFDiffusion in biology illustrate a greater trend: AI is more than a tool for analyzing data; it is becoming an engine for scientific discovery, capable of generating new knowledge, optimizing fundamental processes, and accelerating the pace of scientific progress across numerous fields.

2.2 AI as a Catalyst for Transformation and Future Opportunities

Scientific discovery is a complex process, centered around hypothesis generation, experimentation, data analysis, and the dissemination of findings. Each stage requires creativity, precision, and collaboration, often constrained by the limits of human cognition and resources. AI is transforming this process, acting as a catalyst that accelerates and amplifies the capabilities of researchers, allowing for breakthroughs that were once thought to be out of reach.

2.2.1 Hypothesis Generation

The journey of scientific discovery begins with asking the right questions—formulating hypotheses that guide research. Traditionally, hypothesis generation has been a labor-intensive, iterative process that relies heavily on human intuition and expertise. A critical part of this process involves reviewing existing literature to build a foundation for a current understanding of the field. However, with the exponential growth in scientific publications, the burden of knowledge has become overwhelming. Researchers can only read and process a limited number of papers daily, and their selections may introduce biases due to personal preferences or accessibility. Recent advancements in large language models (LLMs) have transformed the research landscape. For example, with the help of LLMs, researchers can now analyze, extract, and synthesize information from hundreds to thousands of papers in hours (Wu et al., 2025). These tools drastically reduce the time spent on literature reviews, allowing scientists to focus on refining their hypotheses instead of getting stuck in the initial stages of knowledge gathering. LLMs also expand their utility to inventors conducting patent searches, minimizing redundancy and saving resources. By the end of 2024, AI search is emerging as a key application, and in 2025, it is expected to proliferate (Cahn, 2024). There are also emerging trends in AI search engines, with predictions that domain-specific solutions will become widely adopted—for instance, physicians and healthcare providers may use OpenEvidence, investors and analysts may use Perplexity, and lawyers may use Harvey. These advancements promise to make vast amounts of knowledge more accessible, allowing researchers to explore questions that would otherwise remain unexplored. AI tools also enable researchers to navigate vast hypothesis spaces efficiently, providing preliminary feedback that refines hypotheses iteratively. For example, in drug design, researchers traditionally spent months determining the 3D structure of proteins from amino acid sequences to visualize and validate their designs. Tools like AlphaFold revolutionize this hypothesis-generation process, generating highly accurate protein structures within mere seconds. This capability accelerates iterative cycles, which then allow scientists to rapidly test and refine their hypotheses, ultimately improving both the quality and speed of their research.

AI's role in hypothesis generation is exemplified by its ability to identify optimal solutions to problems with vast and complex search spaces. For example, designing a small molecule drug involves exploring 10^{60} potential options, while creating a protein with 400 standard amino acids presents 20^{400} possibilities. AI-driven algorithms help navigate these immense spaces, pinpointing the most promising solutions for further investigation (Agrawal et al., 2024). Similarly, in the development of FAST-PETase, an enzyme engineered to degrade PET plastics, machine learning algorithms analyzed data to predict which mutations of the naturally occurring PETase enzyme were most likely to enhance its efficiency (Lu et al., 2022). AI's assistance in discovering effective solutions to plastic waste degradation showcase its potential in addressing real-world challenges.

Beyond efficiency, AI facilitates the generation of deeper, more impactful hypotheses. As Demis Hassabis, CEO and co-founder of DeepMind, once noted, science can be thought of as a tree of knowledge, where solving fundamental "root node problems," such as protein structure prediction or quantum chemistry, can unlock entirely new branches of research and applications (Hassabis, 2024). By integrating insights from vast datasets and leveraging advanced computational models, AI can help researchers refine their questions and propose bold, foundational hypotheses that push the boundaries of science. These tools enable scientists to shift their focus from incremental improvements to addressing root problems with the potential to propel transformative advancements across various fields.

2.2.2 Experimentation

Once hypotheses are formed, testing them through experimentation is often resource-intensive and time-consuming. For experiments like nuclear fusion, which offers a potential clean and limitless energy source, equipment and facilities are difficult to access. AI can help simulate and accelerate these complex experiments. Researchers from Google partnered with the Swiss Federal Institute of Technology Lausanne to demonstrate how to use a reinforcement learning algorithm to control the shape of plasma in a simulation of a tokamak reactor (Degrave et al., 2022). These approaches could be extended to other experimental facilities, such as particle accelerators, telescope arrays, or gravitational wave detectors (Griffin et al., 2024). By enabling virtual experimentation, AI introduces levels of efficiency which were previously unattainable. A particularly exciting development is the use of AI to bridge dimensional gaps in data, effectively enabling "virtual 3D experimentation" from 2D input. Projects like Sora demonstrate the potential of transforming 2D images and video into dynamic 3D models (Li et al., 2024). This allows for the visualization and analysis of phenomena that would otherwise require complex and expensive physical setups. For instance, Sora's ability to generate scenes like a pirate ship sailing in a coffee cup showcases the potential to explore complex 3D interactions in a virtual environment. This capability works beyond mere visualization; it offers a new avenue for scientific inquiry, allowing researchers to manipulate and interact with these virtual 3D models to test hypotheses and gain insights into real-world phenomena. Furthermore, this ability to project from lower to higher dimensions hints at the possibility of modeling even more complex systems, potentially providing a framework for both exploring and understanding four-dimensional spacetime in the future.

Beyond simulation, AI is increasingly being deployed to automate physical laboratories, accelerating the pace of experimentation. This involves merging AI algorithms with robotic systems capable of performing a wide range of laboratory tasks, from sample preparation and manipulation to data acquisition and analysis. For example, Burger et al. (2020) have developed automated platforms for materials discovery, using AI to design experiments, control robotic synthesis equipment, and analyze the resulting materials' properties. These "self-driving labs" can operate 24/7, generating vast amounts of data and expediting the discovery of new materials with desired characteristics. This automation not only speeds up the experimental process but also reduces human error, leading to more reliable and reproducible results. Furthermore, AI can optimize experimental design in real-time, adapting to incoming data and iteratively refining the experimental parameters to maximize information gain, further enhancing the efficiency of scientific discovery.

2.2.3 Implications for Data and Data Science

Data is the cornerstone of scientific inquiry, providing the empirical foundation for testing hypotheses and building theories. It serves as the evidence crucial to validate or refute hypotheses, driving the iterative cycle of discovery and innovation. Through data, researchers can identify patterns, establish correlations, and draw conclusions that enhance our understanding of the natural world. Given its importance, the impact of AI on data science is a fundamental aspect of scientific advancement.

Evolution of Data Definition. The concept of data has evolved significantly over time. Traditionally, data might have been confined to rows and columns in an Excel spreadsheet. Today, AI technology has expanded our ability to analyze diverse data types, including images, videos, audio recordings, and textual information. For instance, Wikipedia, often regarded as a simple reference website, has been used by organizations like OpenAI as a foundational dataset for training large language models (LLMs) (Brown et al., 2020). Similarly, protein sequence databases, which were originally designed for manual searches, have been transformed into resources for AI-driven research, thus enabling the generation of functional proteins from existing sequences (Madani et al., 2023). As AI continues to advance, both the definition of data and its applications will likely expand, driving progress in unforeseen directions.

Advancements in Data Analysis Methods. The methods used to analyze data have also undergone a transformation. Traditional statistical techniques are well-suited for small, structured datasets but may fall short when applied to large-scale and extra high-dimensional data, such as those encountered in protein structure prediction, image recognition, video generation, and natural language processing. To address these challenges, new model architectures have emerged, including convolutional neural networks (CNNs) (LeCun et al., 1989), and transformers (Vaswani et al., 2017). These models have been pivotal in advancing AI applications in science. For example, in genomics, AI-driven models can predict the effects of genetic mutations on protein function, leading to new insights into disease mechanisms and potential therapeutic targets (Xu et al., 2025). Recent progress in data science has also influenced the field of traditional statistics. Researchers have increasingly recognized the value of predictive models that do not rely on predefined assumptions about data generation. This distinction was highlighted by Breiman (2001) two decades ago as the “two cultures” of statistical modeling.

Scaling Computation for Large Datasets. Traditional statistical methods are not optimized for large-scale GPU computation, which can significantly speed up data analysis. AI-driven approaches, particularly deep learning algorithms, leverage GPU parallel processing to handle massive datasets more efficiently. The importance of scalability in AI has been emphasized in Sutton (2019)’s paper on the “bitter lessons,” which argues for the development of generalizable methods that are applicable to various domains. Scaling laws (Kaplan et al., 2020) continue to highlight the advantages of large-scale computation in improving AI model performance. We anticipate the emergence of more efficient algorithms in the next decade, driven by advancements in GPU and other accelerated computing hardware. These developments will significantly contribute to the progress of data science. An example of this trend is the recent DeepSeek model, which integrates various techniques through an algorithm-architecture co-design approach (Liu et al., 2024).

Data Generation and Synthetic Data. Data generation plays a critical role in scientific research; in particular, when real-world data is either scarce or expensive to obtain. AI has facilitated the creation of synthetic data that closely mimics real-world datasets, which provides researchers with valuable resources for experimentation. For example, in computer vision, generative adversarial networks (GANs) have been used to synthesize images of objects, animals, and scenes, enhancing model training and evaluation (Goodfellow et al., 2014). In fields such as healthcare, data collection is both costly and time-intensive. However, not all data points contribute equally to model performance. Active learning algorithms can help prioritize data collection efforts by selecting the most informative samples. Large scientific experiments generate enormous volumes of data, much of which is noisy, unstructured, or incomplete. AI-driven models are highly effective in sifting through such datasets. For example, experiments at the Large Hadron Collider produce over 100 terabytes of data per second, most of which consists of mere background noise. AI-based anomaly detection methods help identify rare, significant events that would otherwise be overlooked (Collaboration, 2025). In genomics and single-cell biology, AI-powered generative models and self-supervised learning methods enhance live-cell imaging, refine protein-RNA expression analyses, and unveil new biological phenomena (National Science Board, 2020).

Data Preprocessing and Cleaning. Data preprocessing—particularly data cleaning—remains one of the most resource-intensive stages of scientific research, as raw datasets often contain missing values, inconsistent units, batch effects, and instrument-specific artifacts that can compromise downstream analysis. These problems are exacerbated by the lack of standardized data formats across laboratories and platforms, requiring researchers to repeatedly construct cleaning pipelines. AI-driven approaches could offer promising solutions by automating tasks such as anomaly detection, imputation, normalization, record linkage, and metadata extraction. Techniques based on representation learning can infer appropriate corrections directly from data structure, while reinforcement learning systems such as RLClean (Peng et al., 2024) learn adaptive cleaning strategies from expert feedback. LLMs further assist by structuring semi-structured logs or notes into machine-readable formats. However, automated preprocessing may introduce silent errors, struggle with rare edge cases, or obscure provenance, underscoring the continued need for rigorous validation and human oversight even as AI reduces the manual burden of data cleaning.

Facilitating Accessibility in Data Science. Analyzing complex datasets often requires extensive domain expertise, especially in specialized fields such as healthcare. Biostatisticians, for example, typically hold advanced degrees and possess deep expertise in niche areas of statistics. However, researchers in other fields may not always have access to such expertise, limiting their ability to derive meaningful insights from data. AI-driven tools like Copilot offer assistance by recommending suitable analytical models and even generating code, thereby democratizing access to these advanced statistical methods (Araújo et al., 2024).

AI in Data Labeling and Annotation. Data labeling and annotation are essential for training machine learning models but are often constrained by the scarcity of labeled data. The majority of collected data remains unlabeled, and manual annotation—particularly in fields like medical pathology—can be both costly and time-consuming. Active learning methods optimize efficiency by allowing models to selectively choose the most informative data points for annotation (Deng et al., 2023). AI can also assist in automated labeling; for example, in 2022, Google researchers used AI to predict protein functions, leading to the addition of new annotations in publicly available databases such as UniProt, Pfam, and InterPro (Griffin et al., 2024).

Addressing Data Privacy in Scientific Research. Data sharing presents significant ethical and privacy challenges, especially in healthcare. Common de-identification techniques often fail to provide sufficient protection, and differential privacy approaches struggle with multimodal datasets that combine diverse data types, such as MRI images with patient narratives. Generative AI models, such as diffusion-based techniques (Song et al., 2021), offer a promising alternative. By learning the underlying distributions of real-world data, these models generate synthetic datasets that preserve statistical relevance while mitigating privacy risks. This capability allows for secure data-sharing practices, enabling collaborative research while safeguarding sensitive information.

2.2.4 Collaboration and Dissemination

The final step of the scientific process involves sharing discoveries and applying them to benefit society (Sever, 2023). AI reshapes how scientific insights are communicated and used by bridging gaps between fields and democratizing access to knowledge (Dessimoz and Thomas, 2024). LLMs like ChatGPT are changing how researchers engage with complex scientific findings. These models can summarize papers and simplify technical information in ways more accessible to researchers from diverse fields (Leontidis, 2024). For example, a physicist can quickly grasp key insights from biological research, or a climate scientist can incorporate economic data into predictive models without needing deep expertise in economics. By reducing the knowledge burden, AI promotes collaboration and accelerates the exchange of ideas across scientific disciplines (Schäfer, 2023).

AI also bridges the gap between science and policy, addressing a long-standing challenge in knowledge dissemination. Research discoveries are often locked in academic journals, only accessible to field experts and difficult for policymakers or professionals in other fields to digest. This barrier slows the translation of discoveries into actionable policies, often taking years for new findings to influence resource distribution or legislative decisions (NIH Collaboratory, 2025). AI changes this dynamic through its ability to tailor scientific communication to different audiences. For example, LLMs can reframe technical papers into summaries, customized explanations, or multimedia presentations suitable for stakeholders with varying levels of expertise. In the future, AI-powered interactive scientific publications could deliver relevant insights to policymakers, streamlining decision-making and accelerating the adoption of innovations. An example of AI's potential in actionable knowledge dissemination is its use in environmental science. Digital twins, such as virtual simulations of ecosystems, provide policymakers with tools to evaluate the outcomes of climate interventions (Li et al., 2023). These simulations allow decision-makers to experiment with resource distributions and policy choices in a risk-free virtual environment, bridging the gap between research and real-world application.

While science has evolved significantly over the past centuries, the methods of knowledge dissemination have remained largely unchanged. The journal publication process, which was established over 400 years ago, continues to dominate how findings are shared, even despite its limitations in the modern era of rapid innovation. This traditional system often struggles to keep pace with the speed of discovery, particularly in fields like mathematics and statistics, where the peer review and revision process can take years to complete (Aczel et al., 2025). To address this gap, platforms like arXiv have emerged, allowing researchers to share pre-publication manuscripts and bypass lengthy delays (Bengio, 2020). AI and LLM technologies have the potential to transform this outdated model. These tools could enable innovative approaches to disseminating knowledge, like real-time collaborative platforms, automated summarization tools, and interactive digital publications. By utilizing such tools, the scientific community can build towards a pace of knowledge dissemination that matches the speed of discovery, fostering a more connected and responsive research ecosystem (Sever, 2023; Youvan, 2024) which is paving the way for a future where discoveries in science can quickly translate into benefits for society.

AI is not just a tool for automation; it is a vital partner in scientific inquiry, enhancing every stage of the discovery process. From hypothesis generation, where AI analyzes vast datasets to identify potential research questions, to dissemination, where AI-driven platforms tailor scientific communication to diverse audiences, AI significantly boosts human ingenuity and accelerates progress (Khalifa and Albadawy, 2024). In hypothesis generation, AI's ability to process large data volumes helps us uncover hidden patterns and correlations, aiding in formulating and refining hypotheses (O'Brien et al., 2024). For example, AI models can predict the effects of genetic mutations on protein function to provide insights into disease mechanisms and potential therapies. During experimentation, AI optimizes experiment design and execution, maximizing information gain and reducing resource use. In materials science, AI platforms autonomously conduct experiments, analyze results, and refine conditions, speeding up the discovery of new materials (Bai and Zhang, 2025). Beyond experimentation, AI profoundly impacts result analysis and interpretation. Advanced models like deep learning handle complex data and extract insights that guide research directions. In genomics and climate science, where data complexity is high, this capability is indispensable (Alharbi and Rashid, 2022). In dissemination, AI transforms knowledge sharing by making scientific insights accessible to researchers, policymakers, and the public, fostering collaboration and real-world application. The integration of AI into science is transformative, addressing pressing challenges, unlocking new knowledge frontiers, and paving the way for rapid scientific advancements.

3 Challenges and Risks

While AI holds great potential, its widespread adoption introduces significant challenges and risks. Addressing these concerns is critical to ensuring that AI serves as a force for progress while minimizing unintended consequences.

Environmental Costs The environmental impact of AI systems presents a growing challenge at the intersection of technological advancement and sustainability. Training a single model like GPT-3 has been estimated to produce approximately 552 tons of CO2e—equivalent to the annual emissions of 123 gasoline-powered vehicles (Patterson et al., 2021). Furthermore, recent analyses suggest that AI-related electricity use may grow by an order of magnitude over the next several years (de Vries, 2023). This growth signifies the dual nature of AI's energy footprint: while enabling technologies that work to benefit the environment, it simultaneously amplifies environmental costs. Demonstrating both the challenge and its potential solution, Google DeepMind achieved a 40% reduction in data center cooling energy by leveraging AI optimization techniques, showing how AI itself can be a tool for energy efficiency (Jones, 2018). Innovations in efficient computing offer promising paths in minimizing AI's environmental footprint. For example, MIT's "Once-For-All" networks have reduced energy costs by up to 90% compared to traditional neural architecture search methods, pointing toward a future where efficiency and innovation can coexist (Cai et al., 2020). However, with the growing demand for AI, balancing scientific advancements with environmental responsibility remains a formidable challenge. Exploring bold and transformative solutions, the ASCEND feasibility study, led by Thales Alenia Space under the Horizon Europe program, investigated the potential of space-based data centers to meet the EU Green Deal's target of net-zero emissions by 2050 (Thales Alenia Space, 2024). The study revealed that deploying data centers in space could significantly reduce CO_2 emissions and water usage compared to Earth-based facilities, leveraging innovations like reusable launchers and robotic orbital assembly technologies. This initiative not only aims to enhance Europe's digital sovereignty and data security but also projects the deployment of a 1-gigawatt space data center by 2050, contributing to a projected market of 23 gigawatts by 2030. By combining advancements in cloud computing, orbital systems, and eco-friendly data infrastructure, the ASCEND project exemplifies the creative solutions needed to balance AI's potential with the need for sustainability.

Technical Challenges AI models often function as "black boxes," where the mechanisms driving their predictions or decisions are opaque even to their developers. This lack of interpretability raises concerns about reproducibility and trust, especially in critical fields such as medicine, where life-changing decisions depend on reliable results. For example, a misinterpreted model output in drug discovery or clinical diagnostics could lead to harmful consequences (Molnar et al., 2020). The development of explainable AI (XAI) systems is crucial to addressing this issue. XAI focuses on creating models that perform accurately and provide clear, interpretable explanations for their outputs. In fields like healthcare and autonomous systems, these advancements could enhance trust and call for more effective human–AI collaboration.

Recent methodological developments in trustworthy machine learning (ML) attempt to address this challenge from multiple directions. First, on interpretability: beyond classical post-hoc explanations, new techniques embed interpretability into model architecture itself — for example, feature-attention mechanisms for tabular or high-dimensional data (Arik and Pfister, 2021), and the union of feature, data, and component attribution under a unified attribution framework (Zhang et al., 2025). Meanwhile, surveys highlight a recent surge in research on interpretability across modalities (text, image, multimodal), reflecting growing interest beyond traditional structured data (Gao and Guan, 2023). Second, on privacy and data governance: as ML is increasingly used in sensitive domains (e.g., healthcare, finance, drug discovery), privacy–preserving techniques such as Differential Privacy (DP), Federated Learning (FL), secure multi-party computation (SMC) and related privacy-enhancing technologies (PETs) are gaining traction (Mentzas et al., 2024; Zhang et al., 2023). Integrating these with XAI, however, introduces new trade-offs: a recent study shows that applying privacy methods can significantly alter the explanations derived from common attribution methods, making some explanations less reliable (Saifullah et al., 2024). Third, on robustness, fairness, and distributional/generalization issues: trustworthy ML increasingly treats robustness to adversarial attacks, distribution shift, and fairness (e.g., bias mitigation) as co-equal design objectives alongside interpretability and accuracy. For instance, unified frameworks attempt to reason about model behavior under data shifts or adversarial perturbations while preserving interpretability and fairness (Jiang et al., 2023). Empirical studies also investigate the trade-offs among explainability, fairness, privacy, and robustness: improving one dimension (e.g., privacy with DP) may degrade others (e.g., fairness or explanation fidelity) (Kemmerzell and Schreiner, 2024).

Thus, modern "trustworthy ML" frames interpretability, privacy, robustness, and fairness as co-dependent objectives. While this multidimensional view offers a more holistic path toward deploying ML in high-stakes domains (like medicine), it also reveals inherent tensions. Navigating these trade-offs remains a central open challenge. However, as of now, reaching a balance between model performance and explainability remains a complex challenge.

Ethical and Security Risks The ethical implications of AI are vast, ranging from bias in training data to the dual-use nature of its applications, where AI can be implemented for both beneficial and harmful purposes. Many AI systems are trained on datasets that may unintentionally perpetuate societal biases, leading to discriminatory outcomes. For instance, AI-driven hiring tools have been shown to reinforce gender and racial disparities when trained on biased historical data (Hagerty and Rubinov, 2019). The dual-use potential of AI further complicates its ethical deployment. AI systems developed for beneficial purposes, such as drug discovery, can be repurposed for harmful activities. A striking example is the use of AI to design toxic molecules, as demonstrated in a controlled experiment to highlight these risks (Urbina et al., 2022). Such scenarios underscore the urgent need for clear ethical frameworks and governance to prevent the misuse of AI.

Data Challenges and Privacy Concerns The quality and availability of data remain critical bottlenecks for AI development. Noisy, incomplete, or biased datasets can compromise model performance and lead to unreliable outcomes. In addition to this, the growing reliance on sensitive datasets, such as patient health records or financial data, raises significant privacy concerns (Murdoch, 2021). Traditional de-identification methods often fall short, especially when combined datasets can inadvertently re-identify individuals. Emerging solutions, such as differential privacy algorithms and synthetic data generation using generative AI models, offer potential solutions for secure data sharing. However, these approaches are not without limitations. Creating synthetic multimodal datasets that preserve privacy while maintaining utility is an ongoing research challenge (Acosta et al., 2022). Striking a balance between data accessibility and privacy protection is crucial to fostering collaboration without compromising trust (Aldoseri et al., 2023).

Systemic and Societal Risks The adoption of AI into different sectors could worsen existing inequalities. Access to advanced AI tools and infrastructure is often limited to well-funded institutions or private companies, which creates disparities in research capabilities. This unequal access could widen the gap between resource-rich and under-resourced regions, thus limiting the democratization of AI's benefits. Additionally, as AI becomes more embedded in decision-making processes, there is a risk of over-reliance on these systems. Automated decision-making can lead to unintended consequences when human oversight is reduced or when models fail to account for nuanced, context-specific factors. Ensuring that AI systems are used as tools to support rather than replace human judgment is essential for ethical deployment (Galaz et al., 2021).

4 Potential Future of AI

AI's role in science is still evolving, with exciting possibilities on the horizon. We conceptualized its progression through three phases—Keplerian, Edisonian, and Einsteinian—each reflecting a deeper integration of AI into scientific progress. These phases illustrate the growing abilities of AI, from pattern recognition to autonomous experimentation, and ultimately, to a form of Artificial General Intelligence (AGI) capable of hypothesis generation and foundational innovations.

The Keplerian Phase: Data-Driven Discovery The Keplerian phase represents the current era of AI-driven discovery, where AI serves as an advanced analytical tool for identifying patterns and relationships in datasets. Inspired by Johannes Kepler, who synthesized Tycho Brahe's astronomical observations into universal laws of planetary motion, AI in this phase amplifies human intelligence by processing and analyzing massive datasets to uncover insights that would otherwise remain hidden. Modern examples of Keplerian AI include LLMs like GPT-4 that leverage transformer-based architectures (Vaswani et al., 2017) to generate coherent language and summarize scientific findings by predicting the most likely sequence of tokens. Similarly, AlphaFold trained on Protein Data Bank (wwPDB Consortium, 2019), which predicts protein structures with remarkable accuracy based on amino acid sequences, thus revolutionizing structural biology. In astronomy, AI sifts through terabytes of telescope data to detect phenomena like gravitational waves or identify exoplanets. Climate science benefits from AI-driven weather prediction models that integrate high-dimensional meteorological data and outperform traditional systems in short-term forecasting. Despite these advances, Keplerian AI systems are fundamentally limited to training on pre-existing datasets and cannot reason beyond observed patterns. While they significantly amplify our ability to extract knowledge, they remain tools that depend on human-provided inputs and lack the capacity for independent reasoning.

The Edisonian Phase: Experimental Exploration The Edisonian phase signifies a shift toward AI systems that actively conduct and optimize experiments, embodying Thomas Edison's systematic trial-and-error approach to invention. In a recent interview, Richard Sutton, a leading figure in AI, has emphasized the limitations of current AI systems that rely heavily on learning from static data. He advocates for dynamic learning, such as reinforcement learning, which mimics human trial-and-error to refine strategies and solve problems in real time (Silver and Sutton, 2025). Recent advancements in reinforcement learning, such as Reinforcement Learning with Fine-Tuning (ReFT) (Trung et al., 2024), have gained popularity and are now used in conversational models like the latest versions of ChatGPT. These developments mark the transition from static analysis to dynamic experimentation. In this phase, AI goes beyond data analysis to become an experimental collaborator. For instance, "self-driving labs" automate experimental workflows by designing, executing, and analyzing experiments with minimal human intervention. These systems have been crucial in accelerating the discovery of materials like high-entropy alloys by iterating through thousands of experimental conditions in a fraction of the time it would take using traditional methods (Rao et al., 2022). Similarly, virtual laboratories powered by AI simulate environments that are too costly or hazardous to replicate physically, such as nuclear fusion reactors or drug development processes. Reinforcement learning enables these systems to adapt dynamically, refining experimental strategies in real time based on feedback. The Edisonian phase demonstrates how AI can extend human ingenuity by reducing the time, cost, and resources needed for iterative exploration, making it an indispensable partner in scientific discovery.

The Einsteinian Phase: Foundational Innovation The Einsteinian phase, inspired by Albert Einstein, envisions a future where AI achieves the capacity for foundational innovation and the creation of groundbreaking hypotheses. Unlike the Keplerian focus on summarizing large datasets or the Edisonian emphasis on experimentation, the Einsteinian phase draws from small anomalies and unexplored phenomena to ask entirely new questions. Einstein's approach to discovery often centered on phenomena that others overlooked. For instance, he questioned what would happen if he could travel alongside a beam of light, leading to his Special Theory of Relativity. Through thought experiments such as imagining himself in a falling elevator (the equivalence principle) or considering a beam of light bouncing between mirrors inside a moving train to illustrate how time slows down for a fast-moving observer (time dilation), Einstein formulated theories that entirely reshaped our previous understanding of physics and the Universe. Remarkably, these thought experiments were conducted without any direct experimental data (Norton, 1991). His predictions, such as the bending of light by the sun observed during the 1919 solar eclipse and the existence of gravitational waves confirmed in 2016, were validated decades later, highlighting the profound foresight of his theoretical work. In the Einsteinian phase of AI, systems would move beyond data-driven or experimental learning to engage in abstract reasoning and theoretical deduction. As of today, we haven't seen any AI system can achieve this level of intelligence. In the future, people envision these AGI systems could integrate knowledge across multiple fields to propose hypotheses that go beyond current scientific understanding. For example, these systems might formulate unified theories in physics or generate hypotheses about the origins of life, relying on minimal data but instead deep logical and mathematical reasoning.

Together, these three phases—Keplerian, Edisonian, and Einsteinian—capture the growing potential of AI in science. Keplerian AI enhances our ability to analyze data, Edisonian AI automates and optimizes experimentation, and Einsteinian AI pioneers new theoretical models. This progression highlights the increasing autonomy of AI systems while raising questions about the nature of human-AI collaboration in shaping the future of science. As AI transitions from amplifying human intelligence to becoming a creative collaborator, it promises to redefine the boundaries of scientific discovery. Whether uncovering hidden patterns, conducting autonomous experiments, or generating groundbreaking theories, AI holds the potential to lead us all into a new era of understanding and innovation.

5 Conclusion and Call to Action

AI is revolutionizing science, breaking bottlenecks in knowledge synthesis, experimentation, and interdisciplinary collaboration (Zhuang et al., 2020). From decoding protein structures to creating advanced climate models, AI has already demonstrated its great potential across a range of fields. However, its power comes with challenges, including technical limitations, ethical concerns, and environmental costs. Addressing these challenges is imperative to ensure AI's positive and sustainable impact on society.

As AI continues to evolve, its role in science will deepen, not only accelerating progress but also broadening access to knowledge and tools (Vanschoren, 2023). Yet, this future is not guaranteed. Realizing AI's full potential requires responsible development, collective action, and thoughtful integration into the scientific process. Researchers must prioritize the development of explainable and energy-efficient AI systems, promote open science practices, and ensure that ethical considerations remain at the forefront of AI research. By embracing collaboration across scientific fields and addressing AI's environmental footprint, we can unlock its ability to tackle pressing challenges like climate change, global health, and sustainable energy. Beyond research, educators and institutions must equip the next generation to engage with AI by promoting curiosity and interdisciplinary skills early on. Schools and policymakers must support this transition, ensuring equitable access to AI tools and knowledge (Sanusi et al., 2024).

The journey of AI in science mirrors the evolution from Keplerian data analysis to Edisonian experimentation, and ultimately, to Einsteinian innovation. As we stand at this turning point in modern science, it is critical to balance ambition with responsibility. By working together, we can fully harness AI to redefine the boundaries of discovery and drive forth breakthroughs that benefit both humanity and our planet.

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500. https://doi.org/10.1038/s41586-024-07487-w

Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ (2022). Multimodal biomedical AI. *Nature Medicine*, 28(9): 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

Aczel B, Barwich AS, Diekman AB, Fishbach A, Goldstone RL, Gomez P, et al. (2025). The present and future of peer review: Ideas, interventions, and evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 122(5): e2401232121. https://doi.org/10.1073/pnas.2401232121

Agrawal A, McHale J, Oettl A (2024). Artificial intelligence and scientific discovery: A model of prioritized search. *Research Policy*, 53(5): 104989. https://doi.org/10.1016/j.respol.2024.104989

Aldoseri A, Al-Khalifa KN, Hamouda AM (2023). Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences*, 13(12): 7082.

Alharbi WS, Rashid M (2022). A review of deep learning applications in human genomics using next-generation sequencing data. *Human Genomics*, 16(1): 26. https://doi.org/10.1186/s40246-022-00396-x

Araújo R, Ramalhete L, Viegas A, von Rekowski CP , Fonseca TA, Calado CR, et al. (2024). Simplifying data analysis in biomedical research: An automated, user-friendly tool. *Methods and Protocols*, 7(3): 36. https://doi.org/10.3390/mps7030036

Arik SÖ, Pfister T (2021). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8): 6679–6687.

Bai X, Zhang X (2025). Artificial intelligence-powered materials science. *Nano-Micro Letters*, 17(1): 135. https://doi.org/10.1007/s40820-024-01634-8

Bengio Y (2020). Time to rethink the publication process in machine learning. https://yoshuabengio.org/2020/02/26/time-to-rethink-the-publication-process-in-machine-learning/. Blog post. Accessed: 2025-12-10.

Breiman L (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3): 199–215. https://doi.org/10.1214/ss/1009213726

Brown TB, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33: 1877–1901.

Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, et al. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* Apollo – University of Cambridge Repository.

Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, et al. (2020). A mobile robotic chemist. *Nature*, 583(7815): 237–241. https://doi.org/10.1038/s41586-020-2442-2

Burke A (2019). Science and engineering labor force. https://ncses.nsf.gov/pubs/nsb20198 Published by National Science Board (NSB-2019-8). Accessed: 2025-12-10.

Cahn D (2024). AI in 2025: Building blocks firmly in place. https://www.sequoiacap.com/article/ai-in-2025/. Accessed: 2024-12-22.

Cai H, Gan C, Wang T, Zhang Z, Han S (2020). Once-for-all: Train one network and specialize it for efficient deployment. In: *International Conference on Learning Representations (ICLR)*. arXiv preprint, accepted at ICLR 2020.

Collaboration TC (2025). Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Reports on Progress in Physics*, 88(6): 067802.

de Vries A (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10): 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

Degrave J, Felici F, Buchli J, Neunert M, Tracey B, Carpanese F, et al. (2022). Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897): 414–419. https://doi.org/10.1038/s41586-021-04301-9

Deng Y, Yuan Y, Fu H, Qu A (2023). Query-augmented active metric learning. *Journal of the American Statistical Association*, 118(543): 1862–1875. https://doi.org/10.1080/01621459.2021.2019045

Dessimoz C, Thomas PD (2024). AI and the democratization of knowledge. *Scientific Data*, 11(1): 268. https://doi.org/10.1038/s41597-024-03099-1

Díaz-Rodríguez N, Del Ser J, Coeckelbergh M, de Prado ML , Herrera-Viedma E, Herrera F (2023). Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, 99: 101896.

Drew L (2024). Guidelines for academics aim to lessen ethical pitfalls in generative-AI use. *Nature*. https://doi.org/10.1038/d41586-024-01543-1

Fawzi A, Balog M, Huang A, Hubert T, Romera-Paredes B, Barekatain M, et al. (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930): 47–53. https://doi.org/10.1038/s41586-022-05172-4

Galaz V, Centeno MA, Callahan PW, Causevic A, Patterson T, Brass I, et al. (2021). Artificial intelligence, systemic risks, and sustainability. *Technology in Society*, 67: 101741.

Gao L, Guan L (2023). Interpretability of machine learning: Recent advances and future prospects. *IEEE Multimedia*, 30(4): 105–118. https://doi.org/10.1109/MMUL.2023.3272513

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. (2014). Generative adversarial nets. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger K. Q. (eds.), volume 2, NIPS'14, 2672–2680. MIT Press, Cambridge, MA, USA.

Google (2024). AI science forum 2024. https://blog.google/technology/ai/ai-science-forum-2024/. Accessed: 2025-12-10.

Griffin C, Wallace D, Mateos-Garcia J, Schieve H, Kohli P (2024). A new golden age of discovery: Seizing the AI for science opportunity. https://deepmind.google/public-policy/ai-for-science/. Accessed: 2025-12-10.

Hagerty A, Rubinov I (2019). Global AI ethics: A review of the social impacts and ethical implications of artificial intelligence. arXiv preprint: https://arxiv.org/abs/1907.07892.

Hassabis D (2024). Scaling, superhuman AIs, AlphaZero atop LLMs, AlphaFold. https://youtu.be/qTogNUV3CAI. Interview on Dwarkesh Podcast. Accessed: 2025-12-10.

Huang J (2024). Digital biology (keynote address at NVIDIA GTC 2024. Available at: https://blogs.nvidia.com/blog/2024-gtc-keynote/. Accessed: 2025-03-10.

Jiang ZS, Han X, Jin H, Wang G, Chen R, Zou N, et al. (2023). Chasing fairness under distribution shift: A model weight perturbation approach. *Advances in Neural Information Processing*

Systems, 36: 63931–63944.

Jones N (2018). How to stop data centres from gobbling up the world's electricity. *Nature*, 561(7722): 163–166. https://doi.org/10.1038/d41586-018-06610-y

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873): 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaplan J, McCandlish S, Henighan T, Brown TB, Chess B, Child R, et al. (2020). Scaling laws for neural language models. arXiv preprint: https://arxiv.org/abs/2001.08361.

Kellogg Insight (2024). AI is revolutionizing science. are scientists ready? Kellogg Insight, Northwestern University, Available at: https://insight.kellogg.northwestern.edu/article/ai-is-revolutionizing-science-are-scientists-ready. Accessed: 2025-03-10.

Kemmerzell N, Schreiner A (2024). Quantifying the trade-offs between dimensions of trustworthy AI: An empirical study on fairness, explainability, privacy, and robustness. In: *KI 2024: Advances in Artificial Intelligence (LNCS 14434)*, Hotho A, Rudolph S (eds.), 376–390. Springer.

Khalifa M, Albadawy M (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5: 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Lam R, Sanchez-Gonzalez A, Willson M, Wirnsberger P, Fortunato M, Alet F, et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6677): eadi2336. https://doi.org/10.1126/science.adi2336

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4): 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Leontidis G (2024). *Science in the Age of AI: How Artificial Intelligence Is Changing the Nature and Method of Scientific Research.* The Royal Society, London.

Li X, Feng M, Ran Y, Su Y, Liu F, Huang C, et al. (2023). Big data in Earth system science and progress towards a digital twin. *Nature Reviews. Earth & Environment*, 4(5): 319–332. https://doi.org/10.1038/s43017-023-00409-w

Li X, Zhou D, Zhang C, Wei S, Hou Q, Cheng MM (2024). Sora generates videos with stunning geometrical consistency. arXiv preprint: https://arxiv.org/abs/2402.17403.

Liu A, Feng B, Xue B, et al. (DeepSeek-AI Team) (2024). DeepSeek-V3 technical report. arXiv preprint: https://arxiv.org/abs/2412.19437. DeepSeek-AI, December 2024.

Lu H, Diaz DJ, Czarnecki NJ, Zhu C, Kim W, Shroff R, et al. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature*, 604: 662–667. https://doi.org/10.1038/s41586-022-04599-z

MacLeod M, Merz M, Mäki U, Nagatsu M (2019). Investigating interdisciplinary practice: Methodological challenges (introduction). *Perspectives on Science*, 27(4): 545–553. https://doi.org/10.1162/posc_e_00315

Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, et al. (2023). Large language models generate functional protein sequences across diverse families. *Nature Biotechnology*, 41(8): 1099–1106. https://doi.org/10.1038/s41587-022-01618-2

Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6): 1–35. https://doi.org/10.1145/3457607

Mentzas G, Fikardos M, Lepenioti K, Apostolou D (2024). Exploring the landscape of trustworthy artificial intelligence: Status and challenges. *Intelligent Decision Technologies*, 18(2): 837–854. https://doi.org/10.3233/IDT-240366

Molnar C, Casalicchio G, Bischl B (2020). Interpretable machine learning – a brief history, state-of-the-art and challenges. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Koprinska I, Kamp M, Appice A, Loglisci C, Antonie L, et al. (eds.), 417–431. Springer.

Murdoch B (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Medical Ethics*, 22(1): 122. https://doi.org/10.1186/s12910-021-00687-3

National Science Board (2020). Science & engineering indicators 2020: U.S. scientific workforce growth and discovery trends. https://ncses.nsf.gov/pubs/nsb20198. Accessed: 2025-12-10.

NIH Collaboratory (2025). Dissemination to different stakeholder introduction. https://rethinkingclinicaltrials.org/chapters/dissemination/dissemination-different-stakeholders/dissemination-to-different-stakeholder-introduction/. Accessed: 2025-03-10.

Norton JD (1991). Thought experiments in Einstein's work. In: *Thought Experiments in Science and Philosophy* (T Horowitz, GJ Massey, eds.), 129–148. Rowman & Littlefield.

O'Brien T, Stremmel J, Pio-Lopez L, McMillen P, Rasmussen-Ivey C, Levin M (2024). Machine learning for hypothesis generation in biology and medicine: Exploring the latent space of neuroscience and developmental bioelectricity. *Digital Discovery*, 3: 249–263. https://doi.org/10.1039/D3DD00185G

Patterson D, Gonzalez J, Le Q, Liang C, Munguia LM, Rothchild D, et al. (2021). Carbon emissions and large neural network training. arXiv preprint: https://arxiv.org/abs/2104.10350.

Peng J, Shen D, Nie T, Kou Y (2024). RLClean: An unsupervised integrated data cleaning framework based on deep reinforcement learning. *Information Sciences*, 682: 121281. https://doi.org/10.1016/j.ins.2024.121281

Rao Z, Tung PY, Xie R, Wei Y, Zhang H, Ferrari A, et al. (2022). Machine learning–enabled high-entropy alloy discovery. *Science*, 378(6615): 78–85. https://doi.org/10.1126/science.abo4940

Reddy S, Mathur P (2025). Translational AI: Bridging the gap between research and clinical practice. ScienceOpen Preprints. https://doi.org/10.14293/PR2199.001444.v1

Saifullah IT, Mercier D, Lucieri A, Dengel A, Ahmed S (2024). The privacy-explainability trade-off: Unraveling the impacts of differential privacy and federated learning on attribution methods. *Frontiers in Artificial Intelligence*, 7: 1236947. https://doi.org/10.3389/frai.2024.1236947

Sanusi IT, Agbo FJ, Adeleke BS, Osakwe EN, Choi SK, et al. (2024). Stakeholders' insights on artificial intelligence education: Perspectives of teachers, students, and policymakers. *Education and Information Technologies*, 29: 9871–9894.

Schäfer MS (2023). The notorious GPT: Science communication in the age of artificial intelligence. *Journal of Science Communication*, 22(2): Y02. https://doi.org/10.22323/2.22020402

Sever R (2023). Biomedical publishing: Past historic, present continuous, future conditional. *PLoS Biology*, 21(10): e3002234. https://doi.org/10.1371/journal.pbio.3002234

Silver D, Sutton RS (2025). Welcome to the era of experience. *Technical report*, Google DeepMind.

Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B (2021). Score-based generative modeling through stochastic differential equations. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event*. Austria, May 3–7, 2021. OpenReview.net.

Steigerwald E, Ramírez-Castañeda V, Brandt DY, Báldi A, Shapiro JT, Bowker L, et al. (2022).

Overcoming language barriers in academia: Machine translation tools and a vision for a multilingual future. *Bioscience*, 72(10): 988–998. https://doi.org/10.1093/biosci/biac062

Strubell E, Ganesh A, McCallum A (2020). Energy and policy considerations for modern deep learning research. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Conitzer V, Sha F (eds.), volume 34, 13693–13696.

Sutton R (2019). The bitter lesson. http://www.incompleteideas.net/IncIdeas/BitterLesson.html. Blog post. Accessed: 2025-12-10.

Thales Alenia Space (2024). Thales Alenia Space reveals results of ASCEND feasibility study on space data centers. https://www.thalesaleniaspace.com/en/press-releases/thales-alenia-space-reveals-results-ascend-feasibility-study-space-data-centers-0. Accessed: 2025-12-10.

Trung L, Zhang X, Jie Z, Sun P, Jin X, Li H (2024). Reft: Reasoning with reinforced fine-tuning. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Ku L.-W., Martins A, Srikumar V (eds.), 7601–7614.

Urbina F, Lentzos F, Invernizzi C, Ekins S (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, 4(3): 189–191. https://doi.org/10.1038/s42256-022-00465-9

Vanschoren J (2023). Democratising artificial intelligence to accelerate scientific discovery. In: *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*, 224–229. OECD Publishing, Paris.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. (2017). Attention is all you need. In: *Advances in Neural Information Processing Systems*, Guyon I, von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds.), volume 30, 5998–6008.

Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. (2023). De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976): 1089–1100. https://doi.org/10.1038/s41586-023-06415-8

World Economic Forum (2021). Ten challenges to scientific collaboration and how to overcome them. Available at: https://www.weforum.org/agenda/2021/06/ten-challenges-to-scientific-collaboration/. Accessed: 2025-12-10.

Wu S, Ma X, Luo D, Li L, Shi X, Chang X, et al. (2025). Automated literature research and review-generation method based on large language models. *National Science Review*, 12(6): nwaf169. https://doi.org/10.1093/nsr/nwaf169

wwPDB Consortium (2019). Protein data bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1): D520–D528. https://doi.org/10.1093/nar/gky949

Xu W, Li A, Zhao Y, Peng Y (2025). Decoding the effects of mutation on protein interactions using machine learning. *Biophysics Reviews*, 6(1): 011307. https://doi.org/10.1063/5.0249920

Youvan DC (2024). AI-driven democratization of academic publishing: Leveraging preprints for equitable knowledge sharing. https://doi.org/10.13140/RG.2.2.28766.86085

Zhang S, Han T, Bhalla U, Lakkaraju H (2025). Towards unified attribution in explainable AI, data-centric AI, and mechanistic interpretability. arXiv preprint: https://arxiv.org/abs/2501.18887.

Zhang Y, Zeng D, Luo J, Xu Z, King I (2023). A survey of trustworthy federated learning with perspectives on security, robustness and privacy. In: *Companion Proceedings of the ACM Web Conference 2023*, Ding Y, Tang J, Sequeda J, Aroyo L, Castillo C, Houben G.-J. (eds.), 1167–1176.

Zhuang Y, Cai M, Li X, Luo X, Yang Q, Wu F (2020). The next breakthroughs of artificial intelligence: The interdisciplinary nature of AI. *Engineering*, 6(3): 245–247. https://doi.org/10.1016/j.eng.2020.01.009