Open-Access Journal

Review

Volume 1, Issue 1, Year 2026, https://doi.org/10.5281/zenodo.21214021

Toward Trustworthy Artificial Intelligence in Bioinformatics: Ethical Challenges, Algorithmic Bias, and Responsible Innovation

Srishti Sharma[1], Anuja Mishra[2*]

[1]GLA University, Mathura; srishtisharmabt@gmail.com

[2]GLA University, Mathura; anujamishra7777@gmail.com

Received: 01.06.26 | Accepted: 07.06.26 | Published: 12.06.26

Abstract

Current bioinformatics research has been impacted by the rapid advancements in AI and will create faster or more accurate ways to analyze biological data and provide doctors with personalized medicine. In addition, algorithmic bias is a significant issue because the historical inequalities and structural constraints that exist in the training dataset and model evaluation will lead to using AI in a way that perpetuates discrimination. Complex machine learning models are difficult to understand and interpret, which is a problem that arises from the lack of transparency inherent in their design. This review examines ethical, social and trust issues related to artificial intelligence in bioinformatics, with a focus on the root causes of bias throughout the data/modelling lifecycle and their impact. The review also outlines key concepts related to transparency, explainability and accountability as critical tenets of reliable biomedical AI, and discusses emerging standards for responsible usage. Finally, the review discusses future directions for bioinformatics and includes an overview of the need for causal and explainable methods in bioinformatics, the need for inclusive governance structures, and the need for equitable pathways of global innovation in AI and highlights ethical integrity/trust as enablers of long-term, progressive bioinformatics innovation.

Keywords

Ethical AI; bias; trust; bioinformatics; algorithmic fairness; responsible machine learning.

1. Introduction

AI (artificial intelligence) has fundamentally altered the way biology is studied and how precision medicine is practiced. Now, machine learning and deep learning algorithms serve as the basis for many bioinformatics applications, such as genomic variant interpretation, prediction of protein structure, metagenomics, analysis of biomedical images and decision support in clinical settings [1]. In addition, these advances have dramatically increased the ability to analyze large, multifactorial biological datasets and enabled the discovery of complex patterns at rates and levels of detail that are not possible using previous analytical methods. AI/bioinformatics systems have demonstrated great promise in predicting disease risk based on genomic data, encouraging the selection of therapeutic targets using multiple omics, and facilitating various phases throughout drug discovery/development [2]. With the interest in these technologies increasing as they progress from being primarily research-use tools to routine use in clinical settings, they will likely play an important role in the continued evolution of personalized medicine. Recent years have witnessed an unprecedented acceleration in the development of artificial intelligence technologies, particularly with the emergence of deep neural networks, foundation models, generative AI systems, and large language models (LLMs). These advances have transformed bioinformatics from a predominantly data-processing discipline into an intelligent knowledge-discovery framework capable of extracting clinically meaningful insights from highly complex biological datasets [3]. AI-powered approaches are now routinely applied to genomic sequencing, transcriptomics, proteomics, metabolomics, epigenomics, single-cell analyses, and spatial biology, enabling researchers to uncover intricate molecular mechanisms underlying disease progression and therapeutic response. The convergence of AI with high-throughput omics technologies has significantly enhanced the capacity to identify biomarkers, predict disease trajectories, and support precision medicine initiatives across diverse clinical domains [4].

Furthermore, the growing adoption of AI-driven bioinformatics has coincided with an increasing emphasis on data-centric healthcare and personalized medicine. Modern biomedical research generates vast quantities of heterogeneous and multimodal data originating from electronic health records, medical imaging platforms, wearable devices, genomic repositories, and population-scale biobanks [5]. The integration and interpretation of these datasets present substantial analytical challenges that exceed the capabilities of traditional statistical approaches. Consequently, machine learning and deep learning methodologies have emerged as indispensable tools for managing data complexity, facilitating predictive modeling, and enabling real-time clinical decision support. Despite these advancements, the increasing autonomy and complexity of AI systems have intensified concerns regarding accountability, transparency, fairness, reproducibility, and regulatory oversight [6]. The responsible deployment of AI in bioinformatics has therefore become a central topic within both scientific and policy discussions. International regulatory bodies have emphasized the importance of trustworthy and human-centered AI frameworks for healthcare applications [7]. These initiatives advocate for principles such as transparency, explainability, privacy protection, fairness, accountability, and continuous monitoring

throughout the AI lifecycle. As AI systems increasingly influence biomedical research findings, diagnostic pathways, therapeutic recommendations, and public health decision-making, ensuring compliance with ethical and regulatory standards has become essential for maintaining scientific integrity and public confidence [8].

As the use of AI in bioinformatics has increased, so too have the ethical, technical and social issues that will ultimately affect the sustainability and effectiveness of these technologies. The fact that many complex learning systems often lack transparency raises key concerns regarding trustworthiness, clinical translation, regulatory acceptance and professional confidence. Meanwhile, the sensitive nature of both genomic and clinical data introduces unique privacy and security challenges. The issues of algorithmic bias and fairness represent a major obstacle in AI bioinformatics. Bias may arise during data acquisition, curation, model development, validation, and deployment, leading to systematic performance disparities across demographic populations, genetic ancestries, and clinical subgroups [9]. Trust in AI-based bioinformatics can be viewed as a multidimensional construct that reflects technical reliability, transparency, fairness, reproducibility, and compliance with ethical and regulatory principles. Explainable AI (XAI), model documentation, dataset datasheets, privacy-preserving computation, federated learning, secure multiparty computation, and governance frameworks are increasingly recognized as essential components of trustworthy AI ecosystems in biomedical research and healthcare [10].

The convergence of ethical concerns, algorithmic bias, transparency challenges, and trust deficits highlights the urgent need for a comprehensive examination of responsible AI implementation in bioinformatics. Although numerous studies have explored individual aspects of this field—including explainable AI, fairness-aware machine learning, privacy-preserving analytics, and ethical governance—these topics are frequently investigated in isolation. Such fragmented approaches fail to adequately capture the complex interdependencies that exist among ethical principles, technical methodologies, regulatory requirements, and societal expectations. Consequently, a holistic framework is needed to understand how these dimensions collectively influence the development, validation, deployment, and adoption of AI-enabled bioinformatics systems [11]. This review addresses that gap by providing an integrated and interdisciplinary assessment of AI-enabled bioinformatics through three interconnected pillars: (1) ethical principles governing the responsible development and application of AI, (2) sources and mitigation strategies for algorithmic bias and inequity, and (3) technical, institutional, and regulatory mechanisms that foster trust and transparency in AI-driven systems [12]. Drawing upon recent advances in genomics, proteomics, metagenomics, biomedical imaging, precision medicine, and clinical bioinformatics, this review synthesizes current evidence regarding methodological innovations and governance frameworks. Particular attention is given to explainable AI, fairness-aware learning, privacy-preserving technologies, federated learning, model documentation practices, regulatory compliance, and emerging standards for trustworthy AI. By integrating perspectives from computational biology, data science, bioethics, healthcare practice, and policy governance, this review provides a comprehensive roadmap for researchers, clinicians, developers, regulatory authorities, and other stakeholders seeking to

implement AI responsibly within bioinformatics [13]. Ultimately, the goal is to ensure that future AI-enabled bioinformatics systems not only achieve scientific and technological excellence but also uphold principles of equity, transparency, accountability, and societal trust, thereby maximizing their potential to improve biomedical research and patient outcomes.

2. Bias in data source landscapes

Bias in the original source of data has been a major challenge in creating and implementing AI systems that are sufficiently diverse and representative; this is especially true in the case when the data available does not reflect the full range of diversity in the population or dataset being studied. While this has been particularly true in the areas of bioinformatics and other sciences that have made significant investments to ensure that their dataset(s) are "FAIR" (Findable, Accessible, Interoperable, and Reusable), these principles should not be confused with concepts of equity and social inclusion in research enabled by artificial intelligence (AI). There are many different ways to define and operationalize "fairness". However, the overall goal of the FAIR Principles and their related metrics for evaluating data is to provide better ways to find/discover digital objects, to have technical usability, and to have future access to digital objects [14].

Although these practices may enable equitable access and reuse of scholarly data, there is still no guarantee that AI models trained on FAIR-compliant datasets will be free from social, demographic or structural bias, nor that their resulting outputs will be inclusive. Various alternative frameworks have been created to address ethical and social aspects of data science that are unaddressed by FAIR. One example of this is the FACT framework. The FACT framework identifies four core challenges relating to fairness, accuracy, confidentiality, and transparency. It defines fairness in terms of avoiding biased reasoning and unjust outcomes, even in cases where statistically valid results are technically accurate [15]. Another example is FAIR2. The FAIR2 framework expands on the original FAIR paradigm, applying it to social data and analytics by emphasizing the need to properly frame and articulate the context, from which data is collected, as well as to frame and identify any other contextual information necessary to evaluate the data. The additional principles outlined in FAIR2 will help situate data within its historical, cultural, and experiential context, to encourage examination of how various forms of bias may have affected the way data was collected and how data has been interpreted, and to foster engagement with the communities that are represented and/or impacted by the data [16].

The representativeness of the data used to create analytic models is determined by the characteristics and makeup of the data. It is important to note that representativeness is not an inherent or permanent characteristic of a dataset, but rather should be empirically defined in relation to the specific population for which the model will be utilized. In this regard, a dataset can only be considered representative if it adequately reflects the relevant attributes of the target population [17]. There are established methods for determining the representativeness of data, such as evaluating data provenance, sampling strategies and

validity and consistency of measurement procedures. The representativeness of a dataset can be impacted by a number of elements, including: (1) the sources from which the data are collected; (2) the methods by which individuals are selected or recruited; and (3) the methods and protocols used to collect and measure data. Systematic under-representation will occur when some communities encounter continuous impediments to participating in research or accessing clinical and health care services, thereby affecting the representation of those communities within data resources [18]. There are marked differences in the representation of groups and equity of access to data, all due to structural/institutional-based issues, and significantly, the priority and decisions made by funding agencies, and by research leaders have a major impact on what scientific questions are to be addressed and who will be eligible to participate in research or receive quality, equitable healthcare. Decisions about how to carry out the research and develop the structures and methods for participant recruitment will also have an effect on how research questions will be developed, as well as how data will be curated, analyzed, interpreted, and published [19].

3. Algorithmic fairness

In addition to bias associated with the datasets themselves, algorithmic processes used to model those datasets may also create additional sources of bias that may negatively impact the overall fairness and inclusiveness of the model [20]. As mentioned previously, most of the data used for medical AI is derived from the routine practices of clinicians or from legacy studies, and very few datasets have been collected specifically for use in developing AI systems. Therefore, the dataset chosen for model building determines the entire domain of information from which the AI system will learn. While there may be additional stakeholder knowledge or contextual knowledge about the domain of the data that investigators have, such information cannot be learned by an AI system unless it is explicitly defined in the data. This limitation has many implications, including how the data will be structured and represented for training AI models and, therefore, the defining topology of the data space. The definition of a similarity or distance between data points is an important aspect of most machine learning methods. Because of how complicated distance measurements in high-dimensional and heterogeneous biomedical data are to define, versus low-dimensional physical data, defining meaningful distances in high-dimensional data is much more difficult [21].

Experts in a particular field should select the proper distance/similarity function based on their own expertise as well as thoughtful consideration of measurement scales (discretisation), feature encoding or representation, and how the data are represented overall. This influences how the geometry of the data space is structured and affects how well the resulting models will perform. Within the framework of the learning environment, multiple classifications of distance can implicitly determine how the data will be organized topologically, which may limit which models can be learned based on the assumptions about how complex the model should be [22]. Features selected for use in creating the data representation, transformations used on the data, and similarity measures used to build the model likely introduce systematic bias into the prediction provided by the model. For

example, if ethnic, socioeconomic, or other related contextual data is not available or is poorly represented in the similarity function, then it will not be possible for a model to learn to identify patterns related to those variables. As a result, the model will not apply equally across all the population subgroups and will not be able to explain the differences between two people from different ethnic/socioeconomic groups who receive similar interventions [23].

4. Privacy, Security, and Data Sovereignty

There are significant ethical and legal concerns associated with responsible AI use for bioinformatics research due to the ongoing need to protect privacy, secure data, and uphold data sovereignty. Modern AI systems use vast, disparate and often finely granular datasets, and typically, the more data (volume and diversity) that an AI system has access to, the better it will perform. However, this reliance on vast amounts of data also creates many significant ethical and legal issues, including unauthorised access to data, large-scale data breaches and "secondary" misuse of data by outside parties [24]. Although there are many ethical AI principles and guidance, they alone will not ensure that your organisation uses AI responsibly and ethically. Bioinformatics has demonstrated, through established professional standards, fiduciary responsibilities and systems of accountability at institutions, that ethical commitments need to be supported by domain-specific, regulatory, and governance mechanisms that are enforceable to be effective in practice. There have been multiple, large-scale cybersecurity attacks against the health and life-science ecosystems that illustrate the vulnerability of data-driven AI systems. A well-known example is the cyber-attack on Anthem Inc., a major U.S. health insurance provider, in 2015, when personal and demographic information about over 80 million individuals was stolen [25].

In another example of a significant cyber-attack against a health system, a ransomware attack occurred in 2021 across multiple hospitals and health systems (in this case, the Health Service Executive) during which time critical information systems were shut down, and essential clinical services were delayed. Therefore, the vulnerabilities to which AI-based biomedical infrastructure is subject are on par with, if not heightened beyond, the vulnerabilities of other broad-based digital health systems. Additionally, the need for consistent cryptographic protection, ongoing system monitoring, real-time threat detection, and proactive security governance throughout the entire lifecycle of AI represents an ongoing challenge faced by AI-enabled biomedical research, due to the sensitivity of the data involved; the limited or incomplete regulatory coverage applicable to the development of AI; and the risks associated with the changing technology landscape [26]. Even after the removal of personal identifiers from health data, it is still possible to re-identify individuals from that data, especially if it is linked to other data sources. Furthermore, with increasing commercialization and private entities interested in the collection of patient health information, there is an urgent need to establish governance around data stewardship and AI development, as the objectives of commercial entities may not always be aligned with patients' privacy interests, the public's trust, or the long-term societal good [27].

Moreover, international research collaboration and multinational AI development face large compliance burdens because of the varying data sovereignty requirements of nations/regions. The cumulative result of these burdens emphasises the importance of developing integrated legal, technical, and organisational safeguards beyond mere ethical guidance to support the development of trustworthy AI in bioinformatics by ensuring that privacy, security, and sovereignty are core considerations [28].

5. From Principles to Practice: Defining and Evaluating Trustworthy AI

The ethical use of artificial intelligence extends beyond reducing technical and operational risks, and depends on developing and maintaining trust among researchers, clinicians and contributors of data. Trustworthy Artificial Intelligence (AI) is a key requirement in order for AI to be successfully adopted and socially accepted within bioinformatics. Trustworthy AI is typically described using a number of interrelated principles, including: 1) transparency related to how the model works and makes decisions; 2) accountability of those who develop or deploy the model; 3) fairness in both the performance of the model and the outcomes of the model, across groups of individuals; and, 4) respect for patients' autonomy [29]. In this review, we will combine empirical evidence assessing the extent to which current-day AI models implement these principles in practical bioinformatics and biomedical settings, along with regulatory and policy documents which outline best practices for implementing them. All these sources will provide researchers with a basis for evaluating the current maturity of trustworthy AI and developing practical methods to strengthen the ethical and responsible use of AI in bioinformatics research [30].

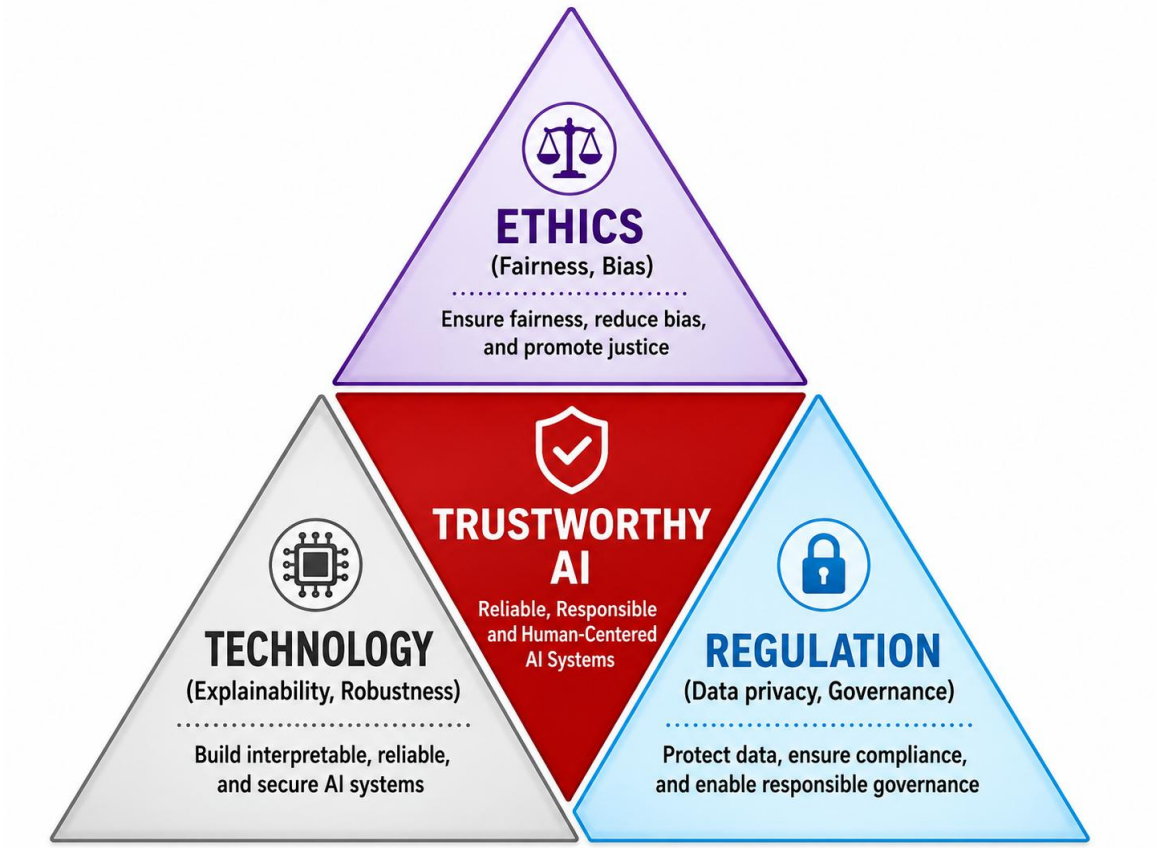

Figure 1. Key components of trustworthy AI: The intersection of ethics, regulations, and technology, emphasizing privacy, fairness, and accountability for ethical AI integration.

Safety, security, and robustness are three necessary attributes of trustworthy artificial intelligence (AI) in the field of bioinformatics, in addition to core values such as trust, transparency (accountability), equity (fairness), and honouring the autonomy of patients. AI systems must be designed and operated in ways that continuously promote the welfare of individual patients and do not create any new or unintended risks when used in either research or clinical settings. Safeguarding AI's infrastructure against cyberattacks, data breaches, and malicious manipulation of models (and/or the data on which the models are built) is also critically important [31]. Lastly, robustness is the ability of an AI system to deliver stable and dependable performance across many different populations, types of data, and settings, and to withstand raucous or adversarial attacks and unforeseen changes in the distributional characteristics of observed data. Moreover, no technical solutions can create trustworthy AI. To accomplish this goal, continued and sustained collaborations among computing scientists, ethical theorists, health care providers, and governmental institutions are critically required to systematically embed safety, security, and robustness into every design, development, and delivery phase of all AI systems [32].

Collaborative interaction among researchers of different fields is critical to balancing the technological ability to produce AI-generated medical decisions with the ethical obligations

of healthcare to ensure that patients are treated fairly and the government standards for patient care are met [33]. For instance, when computer scientists and domain experts work together, they will be able to create algorithms for AI-generated medical decisions that explain themselves to others. However, beyond how algorithmic bias will affect patients in a negative manner, the design and governance of AI systems must also advance the equitable distribution of benefits from AI technologies among many different population groups while being designed and operated under socially responsible guidelines [34]. Systems that are designed and governed without such guidelines are at a high risk of producing algorithmic bias that has a negative impact on patients, particularly those who are historically underserved or socioeconomically disadvantaged. Therefore, it will be imperative to implement bias detection and mitigation strategies at the earliest stage of system design while still preserving security, operational safety, and resiliency within the operational use of bioinformatics and healthcare [35].

5.1. Building Trust through Understanding and Shared Accountability

By implementing sound explainability methodologies and well-defined performance metrics that are reproducible, practitioners can make informed decisions and can be evaluated for regulatory compliance in terms of how well they comply with new transparency requirements that were created to help promote a greater sense of accountability in the use of AI in healthcare and to increase public confidence in the effectiveness and safety of AI-assisted healthcare services [36]. In terms of developing scalable/standardised governance infrastructures (often referred to as “regulatory genomes”) that would facilitate transparency throughout all stages of the AI lifecycle, systematic methodologies for explainability are one way in which researchers can build transparency into all phases of the AI lifecycle as opposed to implementing it after-the-fact (i.e., on an ad hoc basis) [37]. Furthermore, as AI technologies evolve, so too should their mechanisms for providing transparency/interpretable mechanisms so that all high-stakes healthcare decisions made about individuals from various backgrounds can be supported by AI systems that are able to develop rationales for why they provide recommendations based on their respective outputs, be acceptable across different patient demographics and cultures and maintain confidence from different stakeholders in the complexity of the healthcare ecosystem [38].

Implementing successive steps towards achieving measurable objectives of transparency includes defining measurable objectives for transparency in relation to healthcare AI systems, adopting state-of-the-art explanations of those objectives, and establishing accountability mechanisms. These four steps are the way in which the overall goal of achieving trustworthy AI is accomplished with respect to healthcare AI systems. While the internal processes within an organization will improve the reliability and fairness of its services through the application of robust mechanisms, they must also consider administrative rules (e.g., the regulatory requirements) and deal with challenges associated with the delivery of services through the application of healthcare AI technology, such as the fragmentation of services, the privacy of patients, and so on. Collectively, through these combined efforts, we can expect to see the emergence of a more resilient and ethically grounded AI ecosystem that provides transparency and interpretability across patients and providers in diverse settings and diverse clinical environments [39].

5.2. Accountability in AI Decision-Making

When technology like AI is having a greater impact on the decision-making process, especially in both clinical areas and operational areas, it will be necessary to develop clearly defined accountability frameworks that define who is responsible, what they are responsible for, and when in the development and use of an AI tool those responsibilities will apply. According to Michael O'Sullivan, there should be a clear distinction between an autonomous system's accountability (the obligation to provide an explanation or justification for its behaviour), its liability (the legal responsibility for harm caused by the system's actions), and its culpability (whether the autonomous system is morally at fault) [40]. For operational implementation of these distinctions in accountability, Ben Shneiderman suggests establishing multi-level governance of AI tools that cover the various levels of development teams, healthcare organisations, and independent oversight bodies through such means as documentation, traceability, validation of accuracy of the documentation and independent audits. In order for accountability at the policy level in healthcare, Jessica Morley states that there needs to be a direct relationship between accountability for AI systems and the concepts of clinical safety, epistemic validity and fairness so that when AI systems cause harm or exacerbate health inequities, all of the associated developers, providers, and institutions will remain accountable [41].

5.3. Transparency and Explainability: Moving Beyond the Black Box

Developing trustworthy AI solutions in bioinformatics involves far more than simply creating fair, accountable, and private models. It also requires a high degree of robust transparency and explainability. One of the most significant barriers to developing trust in AI solutions is due to the “black box” nature of many existing AI models (especially deep learning models), which severely limits clinicians' ability to understand and evaluate how predictions and recommendations are created. This restriction on interpretability can lead to decreased confidence in clinical practice, increased safety risks, and the inability to adopt AI initiatives in the field [42]. To effectively increase transparency, standardised metrics and validated

interpretability methods must be uniformly included in model development and evaluation processes. In order to evaluate the practical effectiveness of existing bioinformatics AI explainability methods, we have compiled evidence-based studies, regulatory documents, and guidance documents from industry based on PubMed, Scopus, and Web of Science [43]. AI systems must possess two critical qualities: clarity and explainability. These two factors are necessary for the development of trustworthy AI systems. Developers must use models that provide intrinsic interpretability, robust post-hoc explanations, and interfaces that can bridge the operational constraints of the development institution to trustworthy clinical decision support for all users. To achieve the balance of building trust with balance through operational constraints, developers must invest in transparency initiatives [44].

6. Advantages of XAI in Bioinformatics

There are a number of difficulties associated with dealing with large sets of bioinformatics information and these include a lot of variation in terms of their structure, dimensions of data, lack of organisation (for example, no standard format), a lot of background noise or interruptions and missing values, with several different kinds of uncertainty having an impact (such as biases). While bioinformatics is a very data-driven field, the ability to create good machine-learning-based models for solving bioinformatics problems has been limited because there are not many available models that are suitable to handle all the various issues noted above, and also perform well enough [45]. Additionally, there are also significant benefits for both the usability and comprehension of models created using machine learning techniques from having a higher degree of interpretability. Therefore, there is an increased realisation that the learning structure should be interpretable by the user and that even though not every prediction needs a detailed description of why it occurred, there is a fundamental need for at least some form of explanation of the learning process itself [46]. Recent studies have shown that when properly constructed and utilised, AI-based systems that include some form of explanation (e.g. explainable artificial intelligence, or XAI) produce equally consistent performance, reliability and usability compared to traditional artificial intelligence systems, which do not contain any explanation. Additionally, it has been shown that XAI can also assist in developing user trust, provide a systematic means to audit the behaviour of models from an independent party, and enable systematic refinements/iterations of models [47].

6.1. Improves decision fairness and trust

In addition to increasing the potential for AI technologies to have far-reaching impacts (positive or negative) in society and within organisations, fairness has emerged as another critical requirement for high-impact, sensitive applications of AI technologies. Historically, bias or disparate impact has been identified as one of the key barriers to fair and reliable decision-making and has been the focus of extensive research in philosophy and psychology. From a statistical perspective, bias is defined as a systematic difference between actual observed values and the true characteristics of the underlying population from which they are drawn [48]. Within machine learning workflows, bias can be introduced at multiple stages

of the process, including in the data acquisition, sampling, and feature selection, model design and training, hyperparameter optimisation, and interpretation of model outputs to impacted groups (e.g., patients). In particular, there are a number of types of biases in biomedical applications that are embedded in biophysical and clinical datasets and propagated into downstream analytical processes and decision support systems for this population [49]. As statistical generalisation is the foundation upon which machine learning methods operate, machine learning methods can also function unintentionally as statistical discrimination mechanisms, producing systematic advantages for some population groups and producing systematic disadvantages for other population groups through the use of machine learning-based decision-making processes. Therefore, developing explainable and bias-aware machine learning models and methods is critical to achieving fair and trustworthy AI [50].

6.2. Helps avoid practical consequences

Artificial Intelligence (AI) has been extremely helpful in aiding cancer diagnosis and risk stratification, with timely patient diagnosis and accurate classification of patients into high and low-risk categories being essential for clinical management. For instance, when diagnosing breast cancer, AI models use a variety of multi-omic data sources – including genetic abnormalities, copy number variation, gene/microRNA expression and DNA methylation – in addition to imaging studies and clinical records to create more accurate prognostic and diagnostic information [51]. Unfortunately, when deep learning models generate probabilistic predictions for a specific patient, the internal representations of the model are often unclear, making it difficult to determine why a specific diagnosis was made or which molecular or clinical data were given the most weight in making said diagnosis [52]. Interpretable models of machine learning can overcome these limitations by allowing for traceable and transparent reasoning. The local interpretability of a given patient's specific explanation can be determined by identifying either the features that were most influential in producing the individual prediction or the cases, similar to the current one, which contributed the most to the overall decision (global) of the model across all patients. Both interpretations are needed in order for AI-based systems used in cancer diagnosis to be deployed in a responsible manner [53].

6.3. Reduces modeling complexity and improves accuracy

Interpretability not only simplifies model complexity but also enhances its predictive capacity. In genomics-informed cancer evaluation, the key goal is to find biologically informative patterns and establish predictive biological markers that indicate the underlying biological processes and predictor variables for diseases. However, the underlying biology of human beings is determined by complex interactions between thousands of genes rather than just the individual effect of any isolated gene; therefore, chemistries are typically high in dimension (many features/variables) and contain many weakly associated or redundant features/variables, resulting in heavily dispersed low [54]. Clustering coefficient types of data like -omixc. In addition, with cancer clinical trials that involve tens of thousands of genes and relatively few samples the result is a very sparse feature space that results in a greatly

amplified level of noise or error and therefore an increased computational load/cost than for a given number of observed events as well as greater risk of overfitting the model because retained features do not improve prediction accuracy of the model but actually decrease it [55]. Therefore, computationally efficient predictive modelling of cancer requires a process that can methodologically select biologically relevant and highly correlated genes with respect to the outcome of interest (classifying cancer patients) while having the lowest possible redundant feature correlation among genes. Accurately identifying cancer-related genes will increase prediction accuracy, provide a better biological explanation of gene-to-gene associations, and lead to better functionality on the previously identified gene set, resulting in more precise identification of biological factors contributing to the development of cancer. Effective gene ranking will also provide cancer differential markers of predictive reliability, which can further differentiate patients with similar cancer types [56].

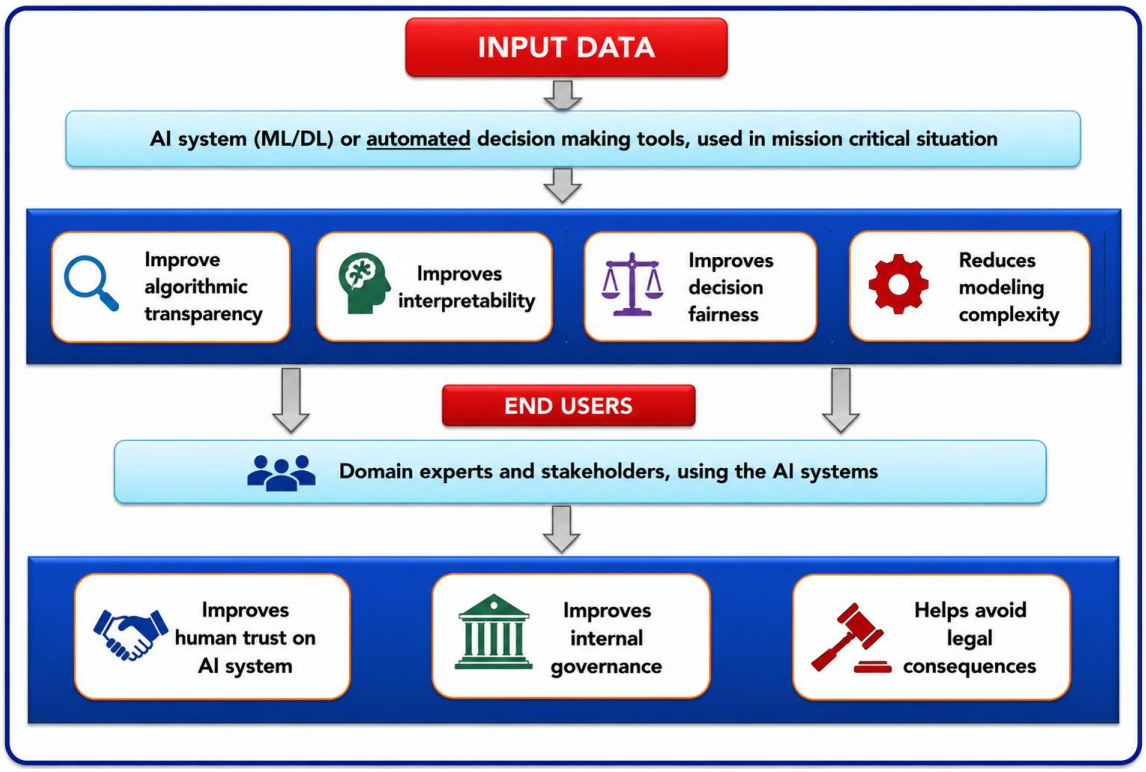

Figure 2. Advantages of XAI in improving interpretability, trust, transparency and fairness in algorithmic decision-making process.

7. AI Governance and Regulation

There is a wider range of entities involved with regulating AI, and there is a move towards creating a regulatory environment that lasts, where the development of advanced AI will be ethical, transparent and socially responsible [57]. All countries, not just the EU, have established core goals for regulating AI that include improved transparency through clarity in the process by which decisions were made, as well as greater accountability for all those that develop or deploy an AI system; reducing algorithmic bias; ensuring fairness; and

implementing stringent processes around data protection. A coordinated, multi-level strategy for regulating AI, utilising both government and local authorities, is important in ensuring that AI development continues to align with generally accepted values and ethical standards [58]. On an international level, the EU has taken a very leading role in setting a legislative and regulatory framework for data protection and trusted AI. The European Commission has proposed numerous regulations concerning the design and deployment of high-risk AI systems, and its High-Level Expert Group on AI has developed some of the most significant ethical principles for the design and deployment of AI systems. In addition to regulatory and ethical frameworks established by the EU, there are a number of ongoing initiatives in the Council of Europe through the work of the Ad Hoc Committee on AI [59]. The focus of these committees is to develop a legally based framework for the design and deployment of AI that provides a basis for human rights, democracy and the rule of law.

Globally, UNESCO, the United Nations' educational, scientific and cultural organisation, has developed a global recommendation document related to the ethical use of artificial intelligence (AI), providing a framework of international normative principles focusing on transparency, accountability, and upholding human rights across all areas of business and life of people [60]. Concerning AI and medicine and biomedical research, the WHO has established guidance on the ethical and human rights aspects of AI for both fields that together establish a framework for implementation based on principles rooted in human autonomy, safety, inclusion, equity, responsibility, explainability, and sustainability. Technical guidance also describes how to implement the above-mentioned principles into actual practice and highlights the importance of building the capacity to implement AI solutions and collaboration between various sectors to support the establishment of a global governance framework governing the use of healthcare-related AI systems [61]. Several federal agencies currently govern and regulate the use of AI in the US. The Federal Trade Commission has created guidelines focused on fairness, transparency, accountability, and avoiding deceptive or biased algorithms. Likewise, the National Institute of Standards and Technology has created guidelines to identify, measure, and mitigate risks associated with the deployment of AI systems. These initiatives will inform broader legislative and policy development, while providing Congress the opportunity to promote responsible government utilisation of AI and protect civil rights and civil liberties while ensuring the protection of both national and economic security [62].

More recently, a comprehensive executive policy framework has consolidated many of these ethical and regulatory priorities into a single, operationally oriented structure. Notably, this framework extends beyond conventional concerns of privacy and trustworthiness by explicitly addressing workforce impacts and establishing concrete mechanisms for implementation and inter-agency coordination [63]. All the different priorities that ethical or regulatory have been condensed into one single Executive Policy Framework recently. This Framework specifically is operational in nature and will not only encompass the traditional privacy and trust issues, but will also provide specific considerations for how they will impact the workforce, and provide clear guidance for implementing these considerations, as well as coordinating with other agencies in order to do so. Important barriers still exist within the

global governance system today, despite the significant amount of work that has already been done to address some of them. For example, regulatory regimes are fragmented across jurisdictions; therefore, an organisation that is operating internationally can find difficulty in complying with conflicting regulations [64]. Technology is evolving rapidly and outpacing the ability to update existing legal or policy instruments, resulting in the need for regulatory models that can adapt more readily and more responsively. In addition, ethical principles tend to be developed at a high level without any operational definitions that are enforceable by law, particularly in the area of algorithmic bias and discrimination and accountability. Lastly, existing legal frameworks favour public policy issues that deal mainly with data protection and privacy while leaving comparatively little room for broader societal issues, including technology-related social challenges such as job loss due to automation, environmental sustainability and long-term risks of deploying systems. Eliminating these shortcomings will require an effort to create an extensive system of international harmonisation of regulations, create a uniform set of ethical standards and develop a more flexible approach to regulating, as well as provide greater legal reinforcement for ethical standards. Such efforts will be critical in order to build a sustainable and ethically sound system of artificial intelligence (AI), especially in key areas like bioinformatics and healthcare [65].

8. Future Perspectives

As the field of AI-assisted bioinformatics continues to evolve (from research to use in practice), several areas will have a high impact on creating bioinformatics technologies that are ethical, fair, and trustworthy in how they function at developing computational systems [66]. These potential areas for collaboration include technical/algorithmic development, methodological improvement, institutional governance structure and linkages between scientific disciplines/cross domains. Overall, this collection of ideas represents a broad support base to address the interrelated issues of ethics, fairness, and trust associated with AI-enabled bioinformatics, as highlighted throughout this review and provides guidance for translating these technologies ethically into the biomedical and healthcare arena [67].

8.1. Fairness by Design and Trustworthy AI Frameworks

Historically, the field of bioinformatics concentrated on modifying the results of their models after the fact with the hiring of experts to mitigate any unfairness. Now, the field is focusing on integrating fairness into the design of their models (i.e., equity and transparency). Fairness by design means that fairness will be included throughout the entire life cycle of a model (e.g. from defining a problem to evaluating it or deploying it) [68]. Future studies should include bioinformatics-specific metrics of fairness that take into account multiple ancestry groups, biological differences between ancestry groups, and clinical risks associated with false positives and false negatives in different ancestries [69]. In addition to developing bioinformatics-specific metrics of fairness, bioinformatics pipelines should also use automatic, ongoing auditing systems to ensure that the model is fair in subsets of the data

and to notify clinicians if their model creates clinically significant differences between groups [70].

8.2. Federated Learning and Differential Privacy for Secure Genomic Data Sharing

The greatest obstacle to large-scale collaborative genomics is how to train a model using data from several different organisations without having access to the actual patient-level data [71]. Future bioinformatics research should prioritise communication-efficient and heterogeneity robust federated learning pipelines that address the varying population structure, sequencing platforms, and demographic imbalances between organisations at the respective sites where they train the model, while also embedding fairness controls for skewed contribution of patient data from various organisations [72]. In conjunction, genome-aware differential privacy techniques must be developed to allow the use of ultra-high-dimensional, highly correlated variances while preserving biologically relevant signals in tasks like gwas, variant calling, and construction of polygenic risk scores. The most efficient approach will be to combine federated learning with differential privacy, where application-specific methods will determine how privacy budgets will be distributed and how much noise will be added, with the aim of maintaining privacy of data versus resulting in a substantial reduction of clinically important performance, and where these methods can be implemented into regulated/consortium-based collaborative bioinformatics processes [73].

8.3. Interdisciplinary Collaboration: Bridging Bioinformatics, Ethics, and Policy

In order to create Trustworthy AI in Bioinformatics, many disciplines need to work collaboratively and develop long-term relationships throughout the lifecycle of research [74]. These collaborators should include bioinformaticians, clinicians, ethicists, legal scholars, policymakers and patient representatives [75]. Future initiatives should emphasise the need for co-designing projects, ensuring that bioinformatics researchers receive training in bioethics, target training on fair machine learning for bioinformatics researchers, and provide regulatory/ethical professionals with technical training that is reciprocal to their respective professions. Institutions can support these initiatives operationally via joint appointments, creating research centres dedicated to Trustworthy AI, and establishing funding streams that will require all teams working on projects related to Trustworthy AI to be comprised of individuals from multiple disciplines. Additionally, it is recommended to establish formal mechanisms for involving patient and advocacy groups in participatory design and evaluation of systems to be used clinically [76].

8.4. Socio-Technical Audits and Bias-Mitigation Tools in Standard Pipelines

It is essential that continuous socio-technical audits are part of the process of moving AI-augmented bioinformatics from being prototypes through to use in reliable clinical settings [77]. All future research regarding bioinformatic software and pipelines should include ongoing audits to allow for systematic evaluation of: (i) how well represented the population data used to train the bioinformatics tools (i.e., the sub-populations contained in that data) are; (ii) the fairness of the algorithms, robustness of the algorithms with respect to

differences in the population and/or datasets on which they were trained, and ability (effectiveness) once introduced into a clinical setting compared to their effectiveness when they were part of the original training dataset; (iii) the impact that algorithms developed for a specific population have had on the quality of clinical care; (iv) whether or not software and/or tools were developed in accordance with all ethical and regulatory standards applicable to the development of the product [78]. Earlier studies provide evaluations pertaining to these issues only at the time of initial product development.

8.5. Emerging Directions and Open Challenges

The six areas identified above are priorities for trustworthy bioinformatics AI, but it is also important to consider future directions. Future directions to be considered include causal inference for the identification of causal relationships vs confounders; hybrid frameworks that integrate mechanistic biological knowledge with machine learning models; and frameworks that identify and define the roles and responsibilities of the stakeholders in AI-assisted decision making, including developers, healthcare institutions, clinicians and regulators [79]. Foundation models, multimodal AI, federated learning, and privacy-preserving computation are among emergent technologies that can enhance the discovery of biomarkers, prediction of diseases and development of drugs and enable the secure sharing of large data sets for collaborative research in biomedical sciences.

Although the opportunities are vast, there are still multiple barriers that remain. These include a lack of availability of diverse and high-quality datasets, algorithmic bias, a lack of model interpretability, a lack of external validation, issues of reproducibility and interoperability, the growing costs associated with the significant computational requirements of advanced AI systems, accessibility, and sustainability, which will all present challenges to the adoption of clinical AI. The identification of a reasonable balance of the use of automation while allowing for meaningful human oversight will be critical in establishing accountability and instilling confidence in AI-assisted decision-making [80]. Future advances will require the development of harmonised regulatory frameworks, standardised evaluation procedures, and continuous monitoring systems that facilitate the deployment of AI technologies in a safe and responsible manner. Ultimately, the long-term success of trustworthy bioinformatics AI depends on embedding fairness, transparency, accountability, privacy, and human-centred design throughout the AI lifecycle, ensuring that biomedical innovations deliver equitable and socially responsible benefits [81].

9. Conclusion

Increasingly powerful artificial intelligence (AI) capabilities are providing bioinformatics researchers with tools for completing tasks at scale that include variant interpretation, protein structure analysis, metagenomics, and clinical decision support. However, the deployment of these capabilities also has significant technical, ethical, and societal risks that are tightly coupled. This review shows how high predictive performance alone does not support clinical or translational use unless the models being utilised are also transparent, fair

across different populations, robust to the effects of real-world data shifts, and provide adequate protection of genomic information that could be used inappropriately. The persistence of algorithmic bias due to imbalanced ancestries, heterogeneous data generation, and mismatches between training and deployment will continue to be a major threat to the equitable application of precision medicine and the reliability of clinical decision support systems. While explainability of AI can provide some form of validation of biological outcomes and add transparency for regulatory purposes, it must be complemented with standards-based documentation, ongoing auditing of performance, and governed by accountable governing bodies. Privacy-preserving computation and responsible data governance will be necessary to facilitate collaboration across large organisations while mitigating the risk of re-identification through genomic data. Overall, trustworthy AI solutions will exist in bioinformatics only when fairness, privacy, accountability, and interdisciplinary collaboration have been integrated as core design principles throughout the life cycle of an AI application, rather than viewed as afterthoughts or regulatory requirements after deployment.

Author Contributions

Conceptualization; AM and SS, Writing; SS, Editing and Final drafting; SS and AM

Funding

This research received no external funding The authors thank the management of GLA University, Mathura for providing an opportunity to complete the review.

Acknowledgments

The authors thank the management of GLA University, Mathura for providing an opportunity to complete the review. The authors declare that there are no conflicts of interest.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

1. Hein, Z. M., Guruparan, D., Okunsai, B., Che Mohd Nassir, C. M. N., Ramli, M. D. C., & Kumar, S. (2025). AI and Machine Learning in Biology: From Genes to Proteins. *Biology*, *14*(10), 1453. **https://doi.org/10.3390/biology14101453**

2. Radha, S. (2025). Bioinformatics driven personalized medicine: integrating multi-omics intelligence for precision therapeutics. *Journal of Modern Techniques in Biology and Allied Sciences*, 14-23. https://doi.org/10.70604/
3. ŞAHiN, E., Arslan, N. N., & Özdemir, D. (2025). Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning. *Neural computing and applications*, *37*(2), 859-965. https://doi.org/10.1007/s00521-024-10437-2
4. Adepoju, D. A., & Adepoju, A. G. (2025). Establishing ethical frameworks for scalable data engineering and governance in AI-driven healthcare systems. *Int. J. Res. Publ. Rev*, *6*(4), 8710-8726. *https://doi.org/10.55248/gengpi.6.0425.1547*
5. Choi, J., Karumbaiah, S., & Matayoshi, J. (2025, March). Bias or insufficient sample size? improving reliable estimation of algorithmic bias for minority groups. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 547-557). https://doi.org/10.1145/3706468.3706540
6. Mundlamuri, R., Gunnam, G. R., Mysari, N. K., & Pujuri, J. (2025). The Evolution of AI: From Classical Machine Learning to Modern Large Language Models. *Ieee Access*. https://doi.org/10.1109/ACCESS.2025.3621344
7. Currie, G., Rohren, E., & Hawk, K. E. (2024). The role of artificial intelligence in supporting person-centred care. In *Person-Centred Care in Radiology* (pp. 343-362). CRC Press. https://doi.org/10.1201/9781003310143
8. Saeidnia, H. R. (2026). Foundations of Artificial Intelligence in Healthcare. *decision-making*, *1*(2), 5. https://doi.org/10.61882/infopub.2026.01.01
9. Yang, E. W., & Velazquez-Villarreal, E. (2025). AI-HOPE: an AI-driven conversational agent for enhanced clinical and genomic data integration in precision medicine research. *Bioinformatics*, *41*(7), btaf359. https://doi.org/10.1093/bioinformatics/btaf359
10. Lilhore, U. K., & Simaiya, S. (2025). Integrating multimodal data fusion for advanced biomedical analysis: a comprehensive review. *Multimodal Data Fusion for Bioinformatics Artificial Intelligence*, 127-145. https://doi.org/10.1002/9781394269969.ch7
11. Wolde, T., Bhardwaj, V., & Pandey, V. (2025). Current bioinformatics tools in precision oncology. *MedComm*, *6*(7), e70243. https://doi.org/10.1002/mco2.70243
12. Ge, S., Sun, S., Xu, H., Cheng, Q., & Ren, Z. (2025). Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective. *Briefings in Bioinformatics*, *26*(2), bbaf136. https://doi.org/10.1093/bib/bbaf136
13. Stephen, O. O., Taiye, O. G., Luka, M. I., Emmanuel, A. M., Agatha, O. A. E., Ipoeminogen, M., ... & Oseni, T. E. (2025). Applications and Effectiveness of Artificial Intelligence in Bioinformatics and Biological Sciences. *African Journal of Advances in Engineering and Technology (AJAET)*, *1*(1), 31-48. https://hdl.handle.net/10520/ejc-aa_ajaet_v1_n1_a3
14. Mishra, A., Sharma, S., & Pandey, S. K. (2025). The present state and potential applications of artificial intelligence in cancer diagnosis and treatment. Recent

Patents on Anti-Cancer Drug Discovery. https://doi.org/10.2174/0115748928361472250123105507

15. Wang, Z., Sannyasi, A., & Jiang, J. (2026). The Diverse and Complex Data Landscape in the Life Sciences Necessitates a Rethinking of Computing and Algorithms. In *Transforming Life Sciences with Cloud Computing and AI: Architecting an Intelligent and Trustworthy Research Ecosystem* (pp. 17-58). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-14222-1_2
16. Goktas, P., & Grzybowski, A. (2025). Shaping the future of healthcare: ethical clinical challenges and pathways to trustworthy AI. *Journal of Clinical Medicine*, *14*(5), 1605. https://doi.org/10.3390/jcm14051605
17. Vetrivel, S. C., Saravanan, T. P., Maheswari, R., & Arun, V. P. (2025). Ethical Considerations Privacy, Fairness, Bias in Genomic Data. In *Applications of Deep Learning in Genomics* (pp. 220-255). CRC Press. https://doi.org/10.1201/9781003558835-12
18. Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., ... & Rashidi, H. H. (2025). Ethical and bias considerations in artificial intelligence/machine learning. *Modern Pathology*, *38*(3), 100686. https://doi.org/10.1016/j.modpat.2024.100686
19. Lyimo, B. (2026). Leveraging Artificial Intelligence to Advance Bioinformatics in Africa: Opportunities, Challenges, and Ethical Considerations in Combating Antimicrobial Resistance. *Bioinformatics and Biology Insights*, *20*, 11779322261427123. https://doi.org/10.1177/11779322261427123
20. Bahangulu, J. K., & Owusu-Berko, L. (2025). Algorithmic bias, data ethics, and governance: Ensuring fairness, transparency and compliance in AI-powered business analytics applications. *World Journal of Advanced Research and Reviews*, *25*(2), 1746-1763. https://doi.org/10.30574/wjarr.2025.25.2.0571
21. Argote, J. G., Maldonado, E., & Maldonado, K. (2025). Algorithmic Bias and Data Justice: ethical challenges in Artificial Intelligence Systems. *EthAIca: Journal of Ethics, AI and Critical Analysis*, (4), 5.
22. Alyami, E. (2026). Ethical Governance of Bioinformatics and Genomic AI Systems: From Compliance to Institutional Legitimacy. *Journal of Medical and Health Studies*, *7*(7), 66-73. https://doi.org/10.32996/jmhs.2027.7.7.8
23. Wu, Y. C., Tuo, N., Shi, G., Li, K., Song, Z., & Li, Y. (2025). The Evolving Role of Artificial Intelligence in Medical Genetics: Advancing Healthcare, Research, and Biosafety Management. *Genes*, *17*(1), 6. https://doi.org/10.3390/genes17010006
24. Sheth, R., & Parekha, C. (2025). Ethical and Privacy Concerns in Integrating Bioinformatics and Cyber-Physical Systems. In *Cyber-Physical Systems Security: A Multi-disciplinary Approach* (pp. 179-190). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-5734-3_8
25. Shaikh, K., Thanki, R., & Shaikh, A. (2026). Regulatory, Ethical, and Data Challenges. In *Artificial Intelligence in Drug Discovery and Development* (pp. 119-137). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-10277-5_6

26. Levy, H. V. (2025). Ethical, legal, and governance dimensions of responsible research and innovation: global perspectives and challenges in emerging technologies. *Legal, and Governance Dimensions of Responsible Research and Innovation: Global Perspectives and Challenges in Emerging Technologies (March 04, 2025)*.

27. Laith, A. E., Jaouni, H., & Mihyar, A. (2025). Addressing cyberbiosecurity challenges in the modern era of biotechnology and artificial intelligence: cyberbiosecurity in the age of biotechnology and AI. *Global Biosecurity*. https://doi.org/10.31646/gbio.282

28. Nkrumah, I. P., Engmann, F., & Adu-Manu, K. S. (2025). Ethical AI in Healthcare: A Comprehensive Review Addressing Privacy, Security, and Fairness. https://doi.org/10.14746/eip.2025.2.2

29. Uddin, M. Z. (2025). Trustworthy AI and explainability. In *Trustworthy Multimodal Intelligent Systems for Independent Living* (pp. 111-137). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97359-8_4

30. Braga, C. M., Serrano, M. A., & Fernández-Medina, E. (2025). Towards a methodology for ethical artificial intelligence system development: A necessary trustworthiness taxonomy. *Expert Systems with Applications*, *286*, 128034. https://doi.org/10.1016/j.eswa.2025.128034

31. Saad, A., Dias, S. B., Alhussein, G., Lyreskog, D., Gerasimou, I., Alves, B., ... & Hadjidimitriou, S. (2026). Clinical AI is Not (Yet) Trustworthy-But It Could Be. *Journal of Medical Internet Research*, *28*, e85433. https://doi.org/10.2196/85433

32. Vo, T. H., & Le, N. Q. K. (2026). Toward trustworthy artificial intelligence in multi-omics: a review of reproducibility, stability, and interpretability. *Briefings in Bioinformatics*, *27*(3), bbag227. https://doi.org/10.1093/bib/bbag227

33. Dixit, M., Kansal, I., Khullar, V., Kumar, R., & Kumar, S. (2025). Analyzing trustworthiness and explainability in artificial intelligence: A comprehensive review. *Recent Advances in Electrical & Electronic Engineering*, *18*(8), 1107-1135. https://doi.org/10.2174/0123520965308169240616144800

34. Kondylakis, H., Osuala, R., Puig-Bosch, X., Lazrak, N., Diaz, O., Kushibar, K., ... & Lekadir, K. (2025). A review of methods for trustworthy AI in medical imaging: The FUTURE-AI Guidelines. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2025.3614546

35. Chander, B., John, C., Warrier, L., & Gopalakrishnan, K. (2025). Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. *ACM Computing Surveys*, *57*(6), 1-49. https://doi.org/10.1145/3675392

36. Emmanuel, U. (2025). BUILDING TRUST IN ARTIFICIAL INTELLIGENCE: ETHICAL CHALLENGES AND TECHNOLOGICAL SOLUTIONS. *Artificial Intelligence (AI)*.

37. Damilola, T. E. (2025). Ethical Considerations and Responsible Artificial Intelligence in Data Practice: A Framework for Trust and Accountability. *Journal of Institutional Research, Big Data Analytics and Innovation*, *1*, 23-37.

38. Deivayanai, V. C., Swaminaathanan, P., Vickram, A. S., Saravanan, A., Bibi, S., Aggarwal, N., ... & Abdel-Daim, M. M. (2025). Transforming healthcare: the impact of artificial intelligence on diagnostics, pharmaceuticals, and ethical considerations–a

comprehensive review. International Journal of Surgery, 111(7), 4666-4693. https://doi.org/10.1097/JS9.0000000000002481

39. Cruz, J. C., Mampuys, R., Reyes, G., & Dryhurst, S. (2025, February). Ethical Implications and Governance Solutions for the Convergence of AI and Synthetic Biology in Healthcare and Agriculture. In *NATO Advanced Research Workshop* (pp. 169-215). Cham: Springer Nature Switzerland.

40. Adebayo, Y., Okunola, C., & Amola, M. (2025). Ethical Implications of Autonomous Decision-Making in Artificial Intelligence: Balancing Innovation and Accountability. *Artificial Intelligence*.

41. Mohammed, S. S. (2025). Navigating Algorithmic Accountability and Ethical Governance in Autonomous Data Analytics Systems: Toward Transparent, Bias-Resistant, and Human-Centric AI Frameworks for Critical Decision-Making. *Bias-Resistant, and Human-Centric AI Frameworks for Critical Decision-Making*. https://doi.org/10.38124/ijisrt/25oct919

42. Pulluri, S. R., Gyani, J., & Pulluri, S. R. (2026). Elucidable Artificial Intelligence in Medical Practice: Enhancing Transparency, Accountability, and Responsible Artificial Intelligence use. In *Explainable AI in Clinical Practice* (pp. 69-89). Academic Press. https://doi.org/10.1016/B978-0-443-44111-0.00004-4

43. Suryadevara, N., Priya, V., & Sharma, S. (2026). From black box to clear box: explainable AI for next-gen pharmacovigilance. *Expert Opinion on Drug Safety*, (just-accepted). https://doi.org/10.1080/14740338.2026.2628822

44. Chinnaraju, A. (2025). Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability. *World Journal of Advanced Engineering Technology and Sciences*, *14*(3), 170-207. https://doi.org/10.30574/wjaets.2025.14.3.0106

45. Budhkar, A., Song, Q., Su, J., & Zhang, X. (2025). Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics. *Computational and Structural Biotechnology Journal*, *27*, 346-359. https://doi.org/10.1016/j.csbj.2024.12.027

46. Patibandla, R. L., Rao, D. M., & Gokul, Y. (2025). Bridging the gap: Understanding genetic discoveries through explainable artificial intelligence. In *Deep Learning in Genetics and Genomics* (pp. 301-311). Academic Press. https://doi.org/10.1016/B978-0-443-27523-4.00021-4

47. Krejcar, O., Abdullah, J., & Namazi, H. (2026). Implementing XAI in Life Sciences: Key Challenges and Pathways to Solutions. *Artificial Intelligence in the Life Sciences*, 100153. https://doi.org/10.1016/j.ailsci.2026.100153

48. George, B., & Wooden, O. S. (2025). Ethical AI and responsible innovation. In *AI Empowered: Pioneering African American Entrepreneurship in the Digital Age* (pp. 105-111). Emerald Publishing Limited. https://doi.org/10.1108/978-1-83662-816-320251012

49. Ponnarengan, H., Rajendran, S., Khalkar, V., Devarajan, G., & Kamaraj, L. (2025). Data-Driven Healthcare: The Role of Computational Methods in Medical Innovation. Computer Modeling in Engineering & Sciences (CMES), 142(1).

50. Goisauf, M., Cano Abadía, M., Akyüz, K., Bobowicz, M., Buyx, A., Colussi, I., ... & Meszaros, J. (2025). Trust, trustworthiness, and the future of medical AI: outcomes of an interdisciplinary expert workshop. *Journal of Medical Internet Research*, *27*, e71236. https://doi.org/10.2196/71236

51. Weiner, E. B., Dankwa-Mullan, I., Nelson, W. A., & Hassanpour, S. (2025). Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. *PLOS digital health*, *4*(4), e0000810 https://doi.org/10.1371/journal.pdig.0000810

52. Ricci, G., Buono, P., Krause, T., Tamla, P., Hemmje, M., De Luzi, F., ... & Mecella, M. (2025, December). Responsible use of AI in Genomics and Ethical Implications. In *2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 7219-7223). IEEE.

53. Mora-Cantallops, M., García-Barriocanal, E., & Sicilia, M. Á. (2024). Trustworthy AI guidelines in biomedical decision-making applications: a scoping review. *Big Data and Cognitive Computing*, *8*(7), 73.. https://doi.org/10.3390/bdcc8070073

54. Izankar, S. V., Kumar, P., & Waghmare, G. (2024, December). Evolution of artificial intelligence in biotechnology: from discovery to ethical and beyond. In *AIP Conference Proceedings* (Vol. 3188, No. 1, p. 080037). AIP Publishing LLC. https://doi.org/10.1063/5.0241106

55. Zhang, J., & Zhang, Z. M. (2023). Ethics and governance of trustworthy medical artificial intelligence. *BMC medical informatics and decision making*, *23*(1), 7. https://doi.org/10.1186/s12911-023-02103-9

56. Boudi, A. L., Boudi, M., Chan, C., Boudi, F. B., & Boudi, A. (2024). Ethical challenges of artificial intelligence in medicine. *Cureus*, *16*(11). https://doi.org/10.7759/cureus.74495

57. Ali, A. (2024). Responsible AI: Aligning Industry Innovation with Ethical Technology Governance.

58. Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert systems*, *41*(7), e13406. https://doi.org/10.1111/exsy.13406

59. Nangoy, A. A., & Chan, P. (2025). Toward Ethical AI: Strategies for Responsible AI Governance. *Journal of Business and Management Studies*, *7*(5), 145-155. https://doi.org/10.32996/jbms.2025.7.5.13

60. Falvo, F. R., & Cannataro, M. (2024, December). Ethics of Artificial Intelligence: challenges, opportunities and future prospects. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 5860-5867). IEEE.

61. Cordeiro, J. V. (2024). Artificial intelligence and precision public health: a balancing act of scientific accuracy, social responsibility, and community engagement. Portuguese Journal of Public Health, 42(1), 1-5. https://doi.org/10.1159/000538141

62. Morris, M. X., Rustom, F., Chun, B., Limon, D., Raghavan, N., Guo, L., & Rajesh, A. (2026). Artificial intelligence in surgical research: transformative impacts and evolving ethical challenges. The American Surgeon™, 92(3), 698-709. https://doi.org/10.1177/00031348251409740

63. Saha, K., & Okmen, N. (2025). Artificial Intelligence in Pharmacovigilance: Leadership for Ethical AI Integration and Human-AI Collaboration in the Pharmaceutical Industry.

64. Singh, B., Kaunert, C., Balusamy, B., & Dhanaraj, R. K. (Eds.). (2025). *Computational Intelligence in Healthcare Law: AI for Ethical Governance and Regulatory Challenges*. CRC Press.

65. Patil, D., Rane, N. L., Desai, P., & Rane, J. (Eds.). (2024). *Trustworthy artificial intelligence in industry and society*. Deep Science Publishing.

66. Sasikumar, S. (2025). The Future of AI Ethics and Responsible Innovation. *Ethics in the Age of AI: Navigating Politics and Security*, 75.

67. Chen, T. (2024). Artificial Intelligence and Drug Design: Future Prospects and Ethical Considerations. *Computational Molecular Biology*, *14*. https://doi.org/10.5376/cmb.2024.14.0002

68. Akhtar, M. A. K., & Kumar, M. (2024). *Towards Ethical and Socially Responsible Explainable AI*. Springer. https://doi.org/10.1007/978-3-031-66489-2

69. Shiferaw, K. B. (2026). *Integrating FAIR Principles and Reporting Guidelines to Support Transparent and Reproducible AI Research in Biomedical Science* (Doctoral dissertation).

70. Usmani, U. A., Usmani, A. Y., & Usmani, M. U. (2023, November). Ensuring trustworthy machine learning: ethical foundations, robust algorithms, and responsible applications. In *2023 International conference on computing, communication, and intelligent systems (ICCCIS)* (pp. 576-583). IEEE. https://doi.org/10.1109/ICCCIS60361.2023.10425285.

71. Mathew, A., & Alex, H. (2025). Federated Learning for Secure Genomic Research: Privacy-Preserving AI Solutions for Precision Medicine. *Science and Technology: Developments and Applications Vol. 9*, 36-43.

72. Matta, S. S., & Bolli, M. (2025). Federated learning for privacy-preserving healthcare data sharing: Enabling global Ai collaboration. *American Journal of Scholarly Research and Innovation*, *4*(01), 320-351. https://doi.org/10.63125/jga18304

73. Mir, B. A., Abbas, S. R., & Lee, S. W. (2026, January). Federated Learning in Healthcare Ethics: A Systematic Review of Privacy-Preserving and Equitable Medical AI. In *Healthcare* (Vol. 14, No. 3, p. 306). MDPI. https://doi.org/10.3390/healthcare14030306

74. Joseph, J. S. (2025). Balancing innovation and biomedical ethics within national institutes of health: Integrative and regulatory reforms for artificial intelligence-driven biotechnology. *Biotechnology law report*, *44*(2), 93-116. https://doi.org/10.1089/blr.2025.360002.jj

75. Lastrucci, A., Pirrera, A., Lepri, G., & Giansanti, D. (2024). Algorethics in healthcare: balancing innovation and integrity in AI development. *Algorithms*, *17*(10), 432. https://doi.org/10.3390/a17100432

76. Patel, A. U., Gu, Q., Esper, R., Maeser, D., & Maeser, N. (2024). The crucial role of interdisciplinary conferences in advancing explainable AI in

healthcare. *BioMedInformatics*, *4*(2), 1363-1383. https://doi.org/10.3390/biomedinformatics4020075

77. Ahmad, A., Vallès, Y., & Idaghdour, Y. (2025). Bias in AI systems: integrating formal and socio-technical approaches. *Frontiers in Big Data*, *8*, 1686452. https://doi.org/10.3389/fdata.2025.1686452

78. Bruneau, G. A. (2025). The Bias Network Approach: a sociotechnical approach to aid AI developers to contextualise and address biases.

79. Obafemi-Ajayi, T., Bright, T. J., Wong, E. F., Wunsch, D., Peckham, J., & Moore, J. H. (2025, June). Ethics vs. Regulation: Converging Frameworks for Trustworthy Human-Centered AI in Biomedical Research. In 2025 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE. https://doi.org/10.1109/IJCNN64981.2025.11227432.

80. Athanasopoulou, K., Michalopoulou, V. I., Scorilas, A., & Adamopoulos, P. G. (2025). Integrating artificial intelligence in next-generation sequencing: advances, challenges, and future directions. *Current Issues in Molecular Biology*, *47*(6), 470. https://doi.org/10.3390/cimb47060470

81. Junaid, M. A. L. (2025). Artificial intelligence driven innovations in biochemistry: A review of emerging research frontiers. *Biomolecules and Biomedicine*, *25*(4), 739. https://doi.org/10.17305/bb.2024.11537