SHAREme.

Scientific Hub of Applied Research in Emerging Medical science & technology

Received: 21.07.2025 Revised: 21.08.2025 Accepted: 26.09.2025

check for updates

Review Article

Ethics, Bias, and Governance: Regulatory Perspectives on AI in Drug Development

Dr. G. Parimala Devi, Kandati Bhavitha*

Pulla Reddy Institute of Pharmacy, Sangareddy-502313, Telangana, India

Abstract: This comparative analysis of AI regulations in key markets provides an overview of the current state of AI regulations in the pharmaceutical industry. The use of AI in drug development is a rapidly evolving field, with regulatory agencies around the world establishing guidelines and regulations to ensure the safe and effective use of AI. The analysis highlights common themes and lessons that can be learned from global regulatory frameworks, including the importance of transparency, data quality, human oversight, and validation. The FDA's framework for AI-related software, the EU's AI regulatory framework, Japan's regulatory sandbox, China's AI-driven healthcare reform, and India's AI-driven healthcare initiatives are discussed as examples of regulatory approaches to AI in drug development. The importance of collaboration, industry-academia partnerships, government support, and public-private partnerships is also emphasized. The analysis concludes that regulatory agencies must stay up-to-date with global developments and best practices to ensure the safe and effective use of AI in drug development.

Keywords: Artificial Intelligence (AI), Drug Development, Ethical AI, Algorithmic Bias, Governance, Pharmacovigilance, Clinical Trials

1. Introduction

The pharmaceutical industry is undergoing a transformative shift with the integration of Artificial Intelligence (AI) into drug development. AI technologies, including machine learning and predictive analytics, are revolutionizing processes such as target identification, clinical trial optimization, and pharmacovigilance, significantly reducing costs and accelerating timelines. However, this rapid adoption raises critical concerns regarding ethical implications, algorithmic bias, and regulatory governance. One of the foremost challenges is ensuring transparency and fairness in AI-driven decision-making. Complex AI models, often referred to as "black-box" systems, lack explainability, making it difficult for regulators and clinicians to validate their outcomes. Additionally, biased training datasets can lead to disparities in drug efficacy and safety across different demographic groups, exacerbating existing healthcare inequities(1).

Regulatory agencies worldwide, including the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), and the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, are actively developing frameworks to oversee AI applications in drug development. These frameworks emphasize data integrity, human oversight, and post-market surveillance, yet gaps remain in standardization and global harmonization.

Correspondence should be addressed to
Kandati Bhavitha; Email: bhavithareddy810@gmail.com

This article explores the ethical dilemmas, sources of bias, and governance challenges associated with AI in drug development. By analyzing case studies and regulatory approaches, we propose strategies to enhance accountability, fairness, and compliance while fostering innovation. The discussion underscores the need for collaborative efforts among policymakers, industry leaders, and AI developers to establish robust, adaptive regulatory frameworks that safeguard public health without stifling technological progress.

2. AI in Drug Development: Applications and Challenges[2]

Key Applications

Artificial Intelligence (AI) is revolutionizing drug development by enhancing efficiency, accuracy, and innovation across the entire lifecycle from early discovery to post-market surveillance. In target identification, AI algorithms process vast genomic, proteomic, and metabolomic datasets to pinpoint promising drug targets with greater precision than traditional methods. Machine learning models can predict protein–ligand interactions, enabling researchers to prioritize targets with the highest therapeutic potential. Within clinical trials, AI optimizes patient recruitment using predictive analytics to ensure diverse and representative study cohorts. It also improves trial design by simulating outcomes, enabling adaptive protocols, and reducing late-stage failures. In pharmacovigilance, AI-powered natural language processing (NLP) tools automatically mine real-world data (RWD) from electronic health records (EHRs), social media, and scientific literature to detect safety signals faster than manual methods. Drug repurposing is another area where AI excels analyzing existing drug databases to uncover new therapeutic uses for approved compounds, thereby saving both time and resources. Finally, in personalized medicine, AI-driven biomarkers and predictive models allow treatments to be tailored to individual patient profiles, improving therapeutic outcomes and minimizing adverse effects.

Benefits

Integrating AI into drug development delivers a range of strategic advantages. In terms of efficiency, AI can screen millions of compounds in a fraction of the time required by conventional methods, reducing development timelines from years to mere months. Accuracy is improved as machine learning models minimize human error in data interpretation, enhancing the reliability of both preclinical and clinical research. AI also promotes cost-effectiveness by refining trial designs and reducing costly late-stage failures, thereby making pharmaceutical R&D more financially sustainable. Additionally, enhanced decision-making through predictive analytics enables researchers to prioritize the most promising drug candidates, while scalability allows for the processing and analysis of massive datasets supporting high-throughput screening and other big-data-driven applications in modern drug development.

Challenges

Despite its transformative potential, AI adoption in drug development faces several obstacles. Data quality remains a critical issue, as AI models depend on complete, unbiased, and representative datasets; poor data can lead to skewed predictions, especially for underrepresented populations. Transparency and explainability pose further challenges, with complex models such as deep learning often functioning as "black boxes" that regulators find difficult to assess. Regulatory gaps also persist, with evolving guidelines and inconsistent requirements across different jurisdictions creating uncertainty for developers. Ethical concerns surrounding patient privacy, informed consent for AI-driven diagnostics, and algorithmic bias add additional complexity. Moreover, integration with existing systems can be problematic, as many pharmaceutical companies still operate with legacy infrastructure that requires costly upgrades to accommodate AI tools. Lastly, a talent shortage in professionals with expertise at the intersection of AI and pharmaceutical sciences limits the pace and breadth of implementation.

3. Ethical Considerations in AI-Driven Drug Development[3,4]

Fundamental Ethical Principles

The application of artificial intelligence in drug development must adhere to core bioethical principles to ensure responsible innovation. Autonomy requires that patients maintain control over their health data and treatment decisions, necessitating clear informed consent processes for AI-driven interventions. Beneficence and non-maleficence demand that AI systems are designed to maximize therapeutic benefits while minimizing potential harms, requiring rigorous validation of algorithms before clinical implementation. Justice principles mandate equitable access to AI-enhanced treatments across different demographic groups, preventing the exacerbation of existing healthcare disparities. These foundational ethics must guide all stages of AI integration, from research design to real-world deployment.

Transparency and Explainability Challenges

The "black box" nature of many advanced AI systems presents significant ethical hurdles in drug development. Complex deep learning models often produce results without clear explanations of their decision-making processes, creating challenges for regulatory review and clinical trust. This opacity becomes particularly problematic when AI influences critical decisions about drug safety, efficacy, or patient eligibility. The ethical imperative for explainability has led to the development of Explainable AI (XAI) techniques, including SHAP values and LIME algorithms, which help decode model decisions. Regulatory agencies increasingly emphasize the need for interpretable AI systems, especially for high-stakes applications like dose optimization or adverse event prediction.

Data Privacy and Informed Consent

The massive data requirements of AI systems raise substantial privacy concerns in pharmaceutical research. Ethical data practices must balance innovation needs with individual rights, particularly regarding sensitive health information. Current frameworks like GDPR in Europe and HIPAA in the U.S. provide important safeguards, but AI applications introduce new complexities. Dynamic consent models are emerging as potential solutions, allowing patients ongoing control over how their data is used in evolving AI systems. Additional ethical challenges include ensuring proper anonymization of datasets, preventing re-identification risks, and establishing clear protocols for secondary data use in AI training processes.

Human Oversight and Accountability

Maintaining appropriate human oversight represents a critical ethical safeguard in AI-driven drug development. While AI can process information faster than human researchers, final decisions about patient care and regulatory approvals must retain meaningful human involvement. This becomes particularly important when AI systems generate novel treatment recommendations or identify off-label uses for existing drugs. Clear accountability frameworks are needed to determine liability for AI-generated decisions, especially when errors occur. Ethical implementation requires well-defined roles for clinicians, researchers, and regulators in supervising AI systems, along with robust audit trails to document human review processes.

Addressing Algorithmic Bias[5]

The potential for embedded biases in AI systems presents one of the most pressing ethical challenges in pharmaceutical applications. Training datasets that underrepresent certain populations can lead to algorithms with reduced accuracy for minority groups, potentially perpetuating healthcare disparities. Ethical AI development requires proactive measures including diverse dataset collection, regular bias testing, and the implementation of fairness metrics during model validation. Post-market surveillance systems must also monitor for emerging biases as AI tools are deployed across different populations. These efforts are essential to ensure equitable benefits from AI advancements in drug development.

Global Equity in AI Benefits

The concentration of AI capabilities in wealthy nations and corporations raises ethical concerns about global access to resulting therapies. Developing countries risk being excluded from both the development and benefits of AI-driven drug discoveries due to technological and economic barriers. Ethical implementation requires international cooperation to build capacity in low-resource settings and establish fair licensing arrangements for AI-discovered treatments. Additionally, ethical frameworks must address potential conflicts when AI systems developed in one cultural context are applied in another, ensuring cultural appropriateness of algorithmically-derived treatments.

4. Bias in AI and Its Implications in Drug Development[6,7]

Understanding Bias in AI

Bias in artificial intelligence (AI) refers to systematic errors in algorithmic decision-making that lead to unfair or discriminatory outcomes. In drug development, AI bias can emerge from multiple sources, including skewed training datasets, flawed model architectures, or human prejudices embedded in data labeling. When AI systems are trained on non-representative clinical trial data often disproportionately including certain demographics they may produce inaccurate predictions for underrepresented groups. This becomes particularly concerning in precision medicine, where biased algorithms could recommend ineffective or unsafe treatments for specific populations. The consequences extend beyond scientific inaccuracy to ethical and legal ramifications, as biased AI may perpetuate healthcare disparities and violate principles of equitable medical care.

Sources of Bias in Drug Development AI

The primary sources of bias in pharmaceutical AI include data bias, where training datasets lack diversity in age, gender, race, or genetic backgrounds; algorithmic bias, arising from improper model design or optimization for narrow populations; and deployment bias, occurring when AI tools are applied to populations different from those in training data. For example, an AI model trained primarily on European genomic data may fail to accurately predict drug responses in Asian or African populations. Additionally, bias can stem from historical inequities in healthcare access, where marginalized groups are underrepresented in medical research. Even well-intentioned AI systems can amplify these biases if not carefully audited, leading to flawed conclusions about drug safety and efficacy across diverse patient groups.

Table 1: Types of Bias in AI Algorithms[2,8]

Type of Bias	Description	Examples
Data Bias	Bias that occurs when the data used to train AI algorithms is incomplete, inaccurate, or biased	Missing data, biased sampling, incomplete data
Algorithmic Bias	Bias that occurs when AI algorithms are designed or implemented in a way that perpetuates existing biases	Biased feature selection, biased model selection, biased hyperparameter tuning
Human Bias	Bias that occurs when humans interpret or use AI-driven results in a way that is biased or discriminatory	Confirmation bias, anchoring bias, availability heuristic
Representation Bias	Bias that occurs when the data used to train AI algorithms does not represent the diversity of the population	Lack of diversity in training data, underrepresentation of certain groups
Measurement Bias	Bias that occurs when the data used to train AI algorithms is measured or collected in a way that is biased or inaccurate	Incorrect or incomplete data collection, biased measurement instruments
Sampling Bias	Bias that occurs when the sample used to train AI algorithms is not representative of the population	Biased sampling methods, incomplete sampling frames

Case Studies of AI Bias in Healthcare[9,10]

Real-world examples highlight the risks of unchecked AI bias in drug development. Boston Medical Center's sepsis detection AI, trained predominantly on data from Caucasian patients, showed significantly lower accuracy for African American individuals, potentially delaying critical care. Similarly, Google's diabetic retinopathy screening tool exhibited reduced performance for patients with darker skin tones due to inadequate representation in training images. In drug discovery, IBM's Watson for Oncology demonstrated bias by recommending more aggressive treatments for younger, privately insured patients, reflecting skewed historical treatment patterns rather than optimal care. These cases underscore the urgent need for bias mitigation strategies to ensure AI-driven drug development benefits all patients equitably.

Case Study 1: Boston Medical Center's Sepsis Detection AI – Racial Bias in Life-Saving Algorithms

Background: Sepsis is a severe, potentially fatal condition caused by the body's overwhelming response to infection, and early detection is critical for survival. To improve patient outcomes, Boston Medical Center (BMC) developed an AI-powered early warning system designed to detect sepsis risk by analyzing electronic health records (EHRs). The system incorporated data such as vital signs, laboratory results, and clinical notes, aiming to identify at-risk patients more quickly than conventional clinical methods. The expectation was that this real-time risk flagging would enable faster intervention, ultimately reducing sepsis-related mortality.

The Bias Problem: While the system demonstrated promising capabilities, its performance revealed significant racial bias, particularly against African American patients. The root cause traced back to training data imbalances, where Caucasian patients were disproportionately represented, leading the model to generalize risk predictions inaccurately for minority groups. Moreover, certain sepsis indicators such as lactate levels and blood pressure tend to vary by race due to both biological and socio-environmental factors. However, the AI was not calibrated to account for these variations. As a result, false negatives occurred more frequently in Black patients, meaning the algorithm failed to flag some cases of sepsis that warranted immediate attention.

Impact: The implications of this bias were severe. Delayed detection meant African American patients faced longer wait times for life-saving treatment, potentially leading to higher morbidity and mortality rates. The issue also amplified existing health disparities, as Black patients already encounter systemic inequities in access to and quality of critical care. Rather than bridging the care gap, the algorithm's design unintentionally reinforced it.

Lessons Learned: This case underscored several critical principles for AI in healthcare. First, diverse and representative training datasets are essential to ensure equitable performance across patient populations. Second, continuous bias auditing must be conducted after deployment to detect and address emerging disparities in algorithmic performance. Finally, mandatory subgroup-specific clinical validation should precede widespread adoption of AI tools, ensuring that predictive accuracy is not disproportionately skewed against any demographic. This case serves as a cautionary example of how even life-saving technology can perpetuate harm if equity considerations are overlooked in design, training, and validation stages.

Case Study 2: Google's Diabetic Retinopathy AI – Skin Tone Bias in Medical Imaging

Background: Diabetic retinopathy (DR) is a major cause of preventable blindness worldwide, and early detection through retinal screening is essential to preserving vision. To address the shortage of ophthalmologists in underserved regions, Google Health developed an AI system capable of analyzing retinal scans to detect DR. The goal was to enable faster, scalable, and more accessible screening, particularly in low-resource settings where specialist access is limited. By automating diagnosis, the technology aimed to reduce delays in treatment and prevent avoidable cases of vision loss.

The Bias Problem: Although the AI demonstrated high overall diagnostic accuracy, performance dropped significantly for patients with darker skin tones. This disparity stemmed largely from dataset imbalance, as the training data lacked adequate representation of individuals across diverse skin pigmentations. Additionally, variations in melanin density can alter the appearance of retinal lesions in imaging, yet the AI had not been calibrated to account for these differences. As a result, both false positives and false negatives increased for Black and South Asian patients either delaying necessary treatment or leading to unnecessary referrals and interventions.

Impact: The bias posed serious clinical and social consequences. Underdiagnosis increased the risk of preventable vision loss in affected groups, while overdiagnosis subjected patients to unneeded treatments, financial burdens, and psychological distress. Beyond individual cases, such disparities threatened to erode trust in AI diagnostics, particularly within communities already skeptical of medical technology due to historical inequities and systemic bias in healthcare systems.

Lessons Learned: This case reinforced the need for diverse and representative datasets in AI training, ensuring balanced inclusion of varied skin tones, ethnicities, and geographic backgrounds. It also highlighted the importance of mandatory algorithmic fairness testing prior to regulatory approval by bodies like the FDA or EMA. Finally, the case underscored that human oversight must remain central to AI-assisted diagnostics ensuring that clinicians can validate, contextualize, and, when necessary, override algorithmic decisions to safeguard accuracy and equity in patient care.

5. Governance and Regulatory Frameworks for AI in Drug Development[11]

The Need for AI-Specific Regulations

The rapid integration of AI into drug development has exposed significant gaps in existing regulatory frameworks. Traditional pharmaceutical regulations were designed for human-led research processes and static algorithms, leaving them ill-equipped to oversee dynamic, self-learning AI systems. This regulatory lag creates uncertainty in areas like algorithm validation, continuous learning systems, and accountability for AI-driven decisions. The current landscape demands new governance models that can ensure patient safety while fostering innovation, addressing unique challenges such as algorithmic transparency, data provenance, and ethical AI deployment across global markets.

Current Global Regulatory Approaches

Major regulatory agencies have adopted divergent strategies for overseeing AI in pharmaceuticals:

- **FDA (U.S.)**: The agency's Digital Health Center of Excellence has pioneered a precertification program for AI/ML-based SaMD (Software as a Medical Device), emphasizing a total product lifecycle approach. Their 2021 AI/ML Action Plan introduced a predetermined change control plan framework for managing algorithm updates.
- **EMA (EU)**: The European Medicines Agency has incorporated AI oversight within its broader digital health strategy, with particular emphasis on GDPR compliance for health data usage. The upcoming EU AI Act will classify certain drug development AI systems as high-risk, requiring stringent conformity assessments.
- **PMDA (Japan)**: Japan's "sandbox" regulatory approach allows temporary real-world testing of AI systems in controlled environments, accelerating innovation while maintaining oversight.
- **NMPA (China)**: China's 2022 guidelines for AI-assisted drug discovery emphasize data security and algorithm transparency, with special provisions for traditional Chinese medicine applications.

Table 2: Fairness Metrics for Evaluating AI Algorithms[12]

Fairness Metric	Description
Demographic Parity	Measures the difference in outcomes between different demographic groups
Equal Opportunity	Measures the difference in true positive rates between different demographic groups
Equalized Odds	Measures the difference in true positive and false positive rates between different demographic groups
Disparate Impact	Measures the difference in selection rates between different demographic groups
Statistical Parity	Measures the difference in outcomes between different demographic groups, adjusted for relevant factors

6. Future Prospects and Recommendations for AI in Drug Development

Strengthening Ethical AI Practices in Pharma

The pharmaceutical industry must prioritize ethical AI governance to maintain public trust and ensure patient safety. AI Ethics Review Boards should be established within all major pharma companies, composed of multidisciplinary teams including bioethicists, data scientists, clinicians, and patient advocates. These boards would evaluate AI projects at all stages - from data collection to deployment - assessing potential biases, privacy risks, and clinical implications. Their oversight should extend beyond initial development to include ongoing monitoring of real-world performance.

Public engagement initiatives are equally critical to demystify AI applications in drug development. Pharmaceutical companies should develop transparent communication strategies explaining how AI is used in research, what safeguards exist, and how patient data is protected. Community advisory panels could provide valuable feedback on AI applications, particularly for vulnerable populations. Educational programs should target healthcare professionals to improve AI literacy, enabling them to better interpret and question AI-driven recommendations in clinical settings.

Policy Recommendations for Responsible AI Adoption

To create a harmonized global framework, the International Council for Harmonisation (ICH) should lead development of universal AI guidelines for pharmaceutical applications. These standards should address key areas including algorithm validation, bias mitigation, and post-market surveillance of AI systems. The guidelines must be flexible enough to accommodate rapid technological advances while maintaining rigorous safety standards across all member nations. Transparency mandates should require comprehensive documentation of AI models submitted for regulatory approval. This includes detailed descriptions of training data demographics, validation methodologies, and performance across different patient subgroups. Regulatory agencies could establish standardized templates for AI model documentation, similar to the Common Technical Document (CTD) format used for drug applications. Open-source repositories for certain non-proprietary AI components could further enhance transparency while protecting intellectual property.

Investment in AI literacy programs is essential to prepare regulators for evaluating increasingly complex AI submissions. Specialized training initiatives should cover fundamental AI concepts, validation methodologies, and emerging risk assessment frameworks. The FDA's Digital Health Center of Excellence could serve as a model for establishing similar training hubs globally. Fellowship programs embedding AI experts within regulatory agencies would facilitate knowledge transfer between industry and regulators.

Implementing Robust Governance Structures[13]

Pharmaceutical companies should adopt proactive AI governance models that anticipate regulatory evolution rather than react to it. This includes establishing internal audit systems that regularly assess AI systems against emerging standards.

Governance frameworks should incorporate explainability-by-design principles, ensuring AI models can provide interpretable outputs without compromising performance. Companies might consider creating Chief AI Officer positions to oversee compliance and ethical implementation across all AI initiatives.

Public-private partnerships can accelerate development of fair and effective AI systems. Collaborative projects between regulators, academic institutions, and pharmaceutical companies could establish benchmark datasets for testing AI algorithms across diverse populations. These partnerships could also develop open-source tools for bias detection and model validation, reducing barriers to compliance for smaller organizations.

Advancing Equity in AI-Driven Drug Development[14]

Future initiatives must prioritize health equity in AI applications. This requires intentional inclusion of diverse populations in both training datasets and clinical validation studies. Regulatory agencies could incentivize diversity through expedited review pathways for AI systems demonstrating robust performance across multiple demographic groups. Global health organizations should establish funding mechanisms to support AI research in underrepresented regions, preventing the concentration of benefits in wealthy nations.

Continuous learning systems should be designed with built-in equity safeguards. This includes real-time monitoring for performance disparities across patient subgroups and automatic alerts when bias thresholds are exceeded. Adaptive trial designs could incorporate AI-driven adjustments to ensure adequate representation of all populations throughout the drug development process.

Table 3: Components of AI Governance Models[15]

Component	Description	Importance
AI Governance Frameworks	Frameworks that provide guidance on the development, deployment, and use of AI	High
AI Ethics Guidelines	Guidelines that provide principles and standards for ensuring AI is used responsibly	High
Data Governance	Policies and procedures for managing and protecting data used in AI	High
Human Oversight	Human review and approval of AI-driven decisions	Medium
Transparency and Explainability	Ability to understand and explain AI-driven decisions	Medium
Accountability	Mechanisms for holding individuals and organizations accountable for AI-driven decisions	High
Continuous Monitoring and Evaluation	Regular review and assessment of AI systems and their impact	Medium

6.5 Preparing for Emerging Technologies[16]

The regulatory framework must remain agile to accommodate next-generation AI technologies like generative AI and quantum machine learning. Special task forces should be established to assess the unique risks and opportunities posed by these advancements. Precompetitive consortia could develop best practices for emerging applications such as AI-designed molecules or synthetic clinical trial data, ensuring safety standards evolve alongside technological capabilities.

International harmonization efforts will be crucial as AI transforms global drug development. Regular summits between major regulatory agencies can align approaches to AI oversight while respecting regional differences in healthcare systems and ethical priorities. The development of mutual recognition agreements for AI validation methodologies could reduce redundant testing and accelerate global access to AI-enhanced therapies.

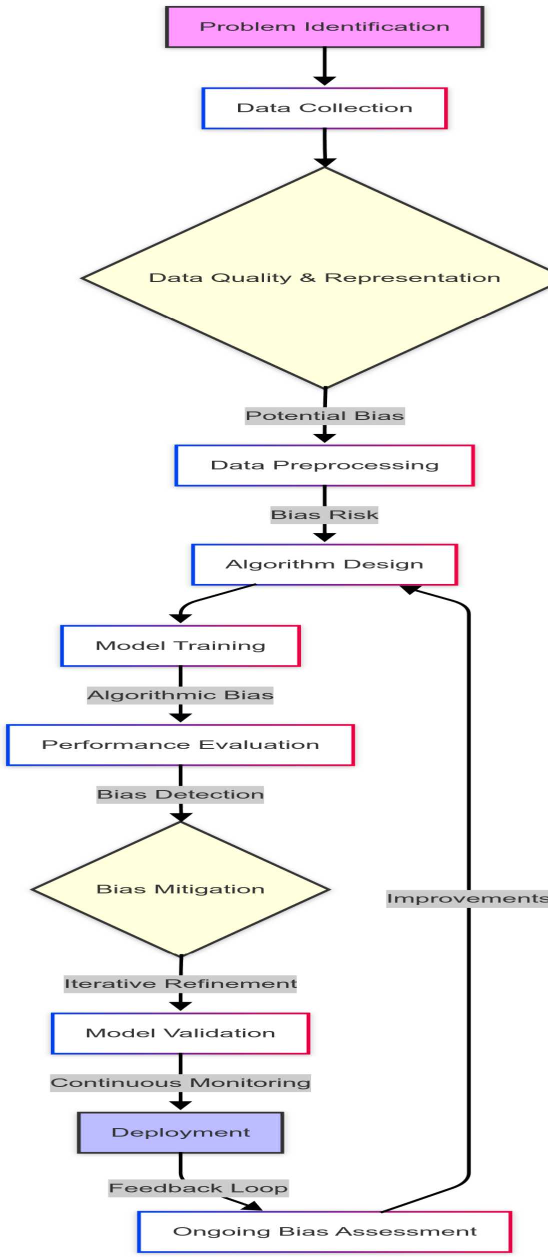

Fig 1: AI Algorithm Development Process

CONCLUSION

Artificial Intelligence (AI) is poised to revolutionize drug development by accelerating discovery, enhancing clinical trials, and improving pharmacovigilance. To fully realize this potential, stakeholders must address ethical, regulatory, and bias-related challenges. Ethical AI must be prioritized, ensuring transparency, accountability, and equitable access while adhering to principles of autonomy, beneficence, and justice. Bias detection and mitigation are essential, as demonstrated by real-world cases like Boston Medical Center's sepsis AI and Google's diabetic retinopathy tool; rigorous fairness audits across diverse populations should be standard practice. Global regulatory harmonization, ideally led by the International Council for Harmonisation (ICH), can streamline approvals while safeguarding patient safety. Transparency and explainability must be mandated, requiring detailed documentation of data, model architecture, and validation protocols in regulatory submissions. Equally important is investment in AI literacy among regulators, clinicians, and pharmaceutical professionals to ensure informed oversight and implementation. Future directions should emphasize developing bias-resistant algorithms, creating standardized fairness metrics, strengthening real-world post-market validation, and fostering public–private partnerships for adaptive governance. By aligning innovation with ethics, robust oversight, and international cooperation, AI can deliver safer, more effective, and equitable therapies worldwide.

REFERENCES

1. Pesapane F, Cuocolo R, Sardanelli F. The Picasso's skepticism on computer science and the dawn of generative AI: questions after the answers to keep "machines-in-the-loop." Eur Radiol Exp. 2024;8(1):81.
2. Skeem JL, Lowenkamp C. Risk, Race, & Recidivism: Predictive Bias and Disparate Impact. Published online June 14, 2016. doi:10.2139/ssrn.2687339
3. Jin Q, Chen F, Zhou Y, et al. Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine. NPJ Digit Med. 2024;7(1):190.
4. Wachter RM, Brynjolfsson E. Will generative artificial intelligence deliver on its promise in health care? JAMA. 2024;331(1):65–69. doi:10.1001/jama.2023.25054
5. Mattu JA, Jeff Larson, Lauren Kirchner, Surya. Machine Bias. ProPublica. Accessed June 21, 2024. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
6. Yaraghi N. Generative AI in health care: opportunities, challenges, and policy. Health Aff Forefront. January 8, 2024. https://www.healthaffairs.org/content/forefront/generative-ai-health-care-opportunities-challenges-and-policy. Accessed April 23, 2024.
7. National Institute of Standards and Technology. Artificial intelligence risk management framework: generative artificial intelligence profile. https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf. Published 2024. Accessed September 25, 2024.
8. Mukerjee A, Biswas R, Deb K, Mathur AP. Multi-objective Evolutionary Algorithms for the Risk-return Trade-off in Bank Loan Management. International Transactions in Operational Research. 2002;9(5):583-597. doi:10.1111/1475-3995.00375
9. Albahri AS, Duhaim AM, Fadhel M, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Inform Fusion. 2023;96:156–191.
10. Entwistle VA, Quick O. Trust in the context of patient safety problems. J Health Organ Manag. 2006;20(5):397–416.
11. Bogen M, Rieke A. Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias. Upturn. 2018. https://apo.org.au/node/210071. Accessed June 21, 2024.
12. Ryan M. In AI we trust: ethics, artificial intelligence, and reliability. Sci Eng Ethics. 2020;26(5):2749–2767. doi:10.1007/s11948-020-00228-y
13. McLennan S, Fiske A, Tigard D, Müller R, Haddadin S, Buyx A. Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Med Ethics. 2022;23(1):6. doi:10.1186/s12910-022-00746-3
14. Stix C. Actionable principles for artificial intelligence policy: three pathways. Sci Eng Ethics. 2021;27(1):15. doi:10.1007/s11948-020-00277-3
15. Kaplan B. Seeing through health information technology: the need for transparency in software, algorithms, data privacy, and regulation. J Law Biosci. 2020;7(1):Isaa062.
16. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. Journal of Global Health. 2019;9(2):020318. doi:10.7189/jogh.09.020318.