AI-Powered Software Testing in Healthcare Applications: A Systematic Review

**Abstract**

The integration of Artificial Intelligence (AI) in healthcare applications has significantly enhanced patient care, diagnostics, and treatment planning. However, ensuring the accuracy, reliability, and safety of these AI-driven healthcare applications is critical, as even minor software defects can have serious consequences, including misdiagnoses, treatment errors, and data breaches. Traditional software testing methods, such as manual testing and rule-based automation, are often inadequate for handling the complexity of modern AI-based healthcare applications. AI-powered software testing has emerged as a solution to address these challenges by leveraging machine learning, deep learning, and natural language processing (NLP) to improve test coverage, automate defect detection, and enhance the robustness of healthcare software systems. AI-driven testing methodologies include automated functional testing, performance testing, security testing, and usability testing, all of which play crucial roles in ensuring software reliability. This systematic review explores AI-powered software testing methodologies, their impact on healthcare applications, and the challenges that need to be addressed for widespread adoption. It also highlights future directions, including the development of explainable AI (XAI) in testing, continuous integration with DevOps, and AI-enhanced real-time validation frameworks to ensure the reliability and security of AI-driven healthcare systems.

***Keywords***

*AI-driven testing, healthcare software, machine learning, automation, validation, safety, regulatory compliance*.

1. **INTRODUCTION**

Healthcare applications, including Electronic Health Records (EHRs), diagnostic tools, and telemedicine platforms, require rigorous software testing to ensure accuracy, security, and compliance with regulatory standards [1]. Traditional testing methods, such as manual and rule-based automated testing, struggle to address the complexity of AI-driven healthcare applications. AI-powered software testing methods leverage machine learning, deep learning, and natural language processing to improve test coverage, detect anomalies, and ensure system robustness [2]. The rapid advancement of Artificial Intelligence (AI) in healthcare has revolutionized patient care, diagnostics, and treatment planning. AI-powered applications now play a crucial role in disease detection, medical imaging, robotic surgeries, telemedicine, and personalized medicine, significantly improving efficiency and accuracy in medical decision-making [3]. However, the increasing reliance on AI-driven healthcare software raises critical concerns about safety, reliability, and compliance with regulatory standards. Even minor defects in software functionality can lead to severe consequences, including incorrect diagnoses, inappropriate treatment recommendations, or data security breaches. Ensuring the robustness and accuracy of healthcare software is paramount, necessitating rigorous software testing methodologies [4].

Traditional software testing approaches, including manual testing and rule-based automation, have been widely used in the healthcare industry [5]. However, these conventional methods often fall short in addressing the complexity and dynamic nature of AI-driven applications. Manual testing is time-consuming, prone to human error, and lacks scalability, while rule-based automation struggles to adapt to AI models that continuously learn and evolve [6]. The unique characteristics of AI-powered healthcare applications, such as self-learning algorithms, real-time data processing, and adaptive decision-making, require advanced testing techniques beyond traditional software validation approaches [7]. AI-powered software testing has emerged as a transformative solution to these challenges by leveraging machine learning, deep learning, and natural language processing (NLP) to enhance the efficiency and effectiveness of healthcare software testing. AI-driven testing methodologies can intelligently identify defects, optimize test cases, and improve test coverage by analyzing large datasets and detecting patterns that traditional testing methods might overlook. Machine learning algorithms can predict potential software failures by analyzing past defects, enabling proactive error detection and mitigation [8]. NLP-based testing is particularly beneficial for evaluating AI-driven clinical decision support systems, medical chatbots, and electronic health record (EHR) systems, ensuring they provide accurate, relevant, and coherent responses [9]. Reinforcement learning is another AI-driven testing technique that dynamically refines test scenarios based on system responses, making it useful for adaptive healthcare applications that continuously learn from new data. Additionally, AI-enhanced fuzz testing introduces unexpected and random inputs to stress-test healthcare applications for vulnerabilities, improving their resilience against cyber threats and system failures. AI-powered regression testing ensures that software updates or modifications do not introduce new defects, which is crucial for maintaining the stability and reliability of healthcare applications over time [10].

The benefits of AI-powered software testing in healthcare are numerous. It improves testing efficiency by automating repetitive test cases, reducing human intervention, and accelerating the validation process. AI-driven testing enhances accuracy and consistency, minimizing human errors and ensuring thorough software validation. Additionally, AI-powered testing is scalable and adaptive, allowing healthcare applications to be continuously tested and refined in response to real-world data and user interactions. This capability is particularly valuable for AI-driven diagnostic tools, which must evolve based on emerging medical research and clinical guidelines [11].

Despite its potential, AI-powered software testing in healthcare also presents significant challenges. One major concern is data privacy and security, as AI-based testing often requires access to sensitive patient data, raising ethical and regulatory issues. Compliance with stringent healthcare regulations such as HIPAA (Health Insurance Portability and Accountability Act) in the United States and GDPR (General Data Protection Regulation) in the European Union is a critical challenge in implementing AI-driven testing frameworks. Furthermore, the high initial investment required for AI-based testing tools and the need for specialized expertise pose additional barriers to widespread adoption [12]. This systematic review aims to evaluate the existing literature on AI-powered software testing for healthcare applications, focusing on key methodologies, benefits, challenges, and regulatory considerations.

1. **Methodology**

To conduct this systematic review, a comprehensive search strategy was implemented to identify relevant studies on AI-powered software testing in healthcare applications. The literature search was conducted across multiple scientific databases, including **PubMed, IEEE Xplore, ACM Digital Library, and Google Scholar,** using specific keywords and Boolean search operatorsto ensure the inclusion of a broad range of studies. Keywords such as **"AI-driven software testing," "machine learning in software testing," "automated software validation," and "healthcare AI systems"** were used in various combinations to retrieve relevant articles. To ensure the relevance and quality of the studies included in this review, specific inclusion and exclusion criteria were applied. Only peer-reviewed articles published between **2015 and 2024** in English were considered, focusing exclusively on software testing methodologies that leverage AI in the context of healthcare applications. Studies that addressed manual software testing without any AI integration or those that were not directly related to healthcare were excluded.

After filtering the relevant articles, data extraction was performed to analyze various aspects of AI-powered software testing in healthcare. The extracted information was categorized based on **the types of AI techniques used, testing approaches, regulatory compliance considerations, challenges, and future research directions.** AI-driven testing methodologies were examined for their impact on software reliability, security, performance, and usability in healthcare applications. Furthermore, particular attention was given to studies discussing AI-powered test automation frameworks, reinforcement learning for test case generation, and natural language processing (NLP) for validating AI-based medical chatbots and decision-support systems.

The review also considered regulatory compliance frameworks such as **HIPAA (USA), GDPR (EU), and FDA regulations**, evaluating how AI-driven software testing aligns with these standards. Finally, challenges and limitations, including ethical concerns, data privacy issues, AI model biases, and the lack of standardized AI testing frameworks, were analyzed. The insights gathered from this systematic review provide a comprehensive understanding of the current landscape, highlighting the potential and limitations of AI-powered software testing in healthcare applications while outlining key areas for future research and improvement.

* 1. **Search Strategy**

A comprehensive literature search was conducted across major scientific databases, including PubMed, IEEE Xplore, ACM Digital Library, and Google Scholar, using the following search terms, AI-driven software testing" AND "healthcare applications, Machine learning in software testing" AND "medical software and Automated software validation" AND "healthcare AI systems.

* 1. **List 1 : Inclusion and Exclusion Criteria**

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Inclusion** | **Exclusion** |
| Publication Year | 2015-2024 | Before 2015 |
| Language | English | Non-English |
| Application Area | Healthcare software testing | Non-healthcare applications |
| Methodology | AI-based software testing | Manual testing without AI |

* 1. **Data Extraction and Analysis**

Following the identification and selection of relevant studies, a structured data extraction process was conducted to systematically analyze the role of AI-powered software testing in healthcare applications. The primary objective of data extraction was to identify and categorize key elements related to AI-driven testing methodologies, their applications, benefits, challenges, and compliance with healthcare regulations. To ensure a thorough and objective analysis, a standardized data extraction template was developed, which included the following key parameters: **publication details (authors, year, journal/conference), AI techniques used, software testing approaches, specific healthcare applications, performance metrics, regulatory compliance considerations, challenges, and future recommendations.**

Each selected study was carefully reviewed to extract information regarding the **types of AI techniques** employed in software testing. These included **machine learning (ML)-based defect prediction, deep learning models for automated test case generation, reinforcement learning for adaptive testing, natural language processing (NLP) for text-based validation, and AI-driven fuzz testing for vulnerability detection.** The extracted data provided insights into how different AImethodologies enhance various testing approaches, such **as functional testing, security testing, performance testing, usability testing, and regression testing.** Additionally, studies wereexamined to determine the level of **automation in software validation,** assessing whether AI wasused for **self-learning test case optimization, anomaly detection, or continuous testing integration** within DevOps frameworks**.**

1. **Results**

The systematic review of AI-powered software testing in healthcare applications revealed several key findings regarding the effectiveness, challenges, and future potential of AI-driven testing methodologies. The studies analyzed demonstrated that machine learning (ML), deep learning, and natural language processing (NLP) significantly enhance the accuracy, efficiency, and automation of software testing processes in healthcare applications [13-16]. AI-powered testing approaches, including defect prediction, automated test case generation, anomaly detection, and security validation, have been successfully implemented in healthcare software systems such as electronic health records (EHR), clinical decision support systems (CDSS), AI-driven diagnostic tools, and telemedicine platforms.

* 1. **Effectiveness of AI-Powered Software Testing**

The studies reported higher test coverage, faster execution times, and improved defect detection rates compared to traditional manual and rule-based automated testing approaches. Machine learning models demonstrated over 90% accuracy [17] in predicting software defects before deployment, reducing the likelihood of critical failures in real-world healthcare applications. Deep learning-based test automation frameworks significantly reduced the need for manual intervention, improving testing efficiency by up to 70% in complex AI-driven applications [18]. Reinforcement learning (RL) was particularly effective in self-learning test case generation, optimizing test coverage while minimizing redundant test cases. Additionally, AI-driven fuzz testing helped uncover previously undetected vulnerabilities in medical software by introducing dynamic, unpredictable test inputs [19].



Figure Effectiveness of AI-Powered Software Testing [20]

* 1. **AI in Different Software Testing Domains**

The studies categorized AI-driven testing across multiple testing domains in healthcare applications:

* Functional Testing: AI-powered testing tools were able to detect logic errors in healthcare decision-support systems with a 25–40% increase in accuracy compared to traditional testing [21].
* Security Testing: AI-driven penetration testing identified previously unknown vulnerabilities in patient data security protocols, ensuring compliance with HIPAA and GDPR regulations [22].
* Performance Testing: Automated AI performance monitoring tools provided real-time optimization of medical software response times, reducing system crashes under high loads by 60% [23].
* Usability Testing: NLP-based AI models were effective in assessing the accuracy and reliability of medical chatbots and virtual assistants, improving their coherence and response accuracy by 30–50% [24].
* Regression Testing: AI-enhanced regression testing frameworks ensured that new updates in EHR systems and telemedicine platforms did not introduce unintended software defects, reducing post-update failures by 40–50% [25].



Figure AI in Different Software Testing Domains

* 1. **AI’s Role in Ensuring Regulatory Compliance**

Several studies highlighted the importance of AI in ensuring compliance with regulatory standards such as HIPAA (USA), GDPR (EU), and FDA (USA). AI-driven testing tools helped automate compliance validation, reducing human error in security and privacy assessments. Automated compliance testing using AI was able to detect non-compliant software behaviors in 15–30% of cases, which may have been overlooked in manual testing. AI also assisted in maintaining audit trails for traceability and accountability in healthcare software development [26].



Figure AI’s Role in Ensuring Regulatory Compliance

1. **Challenges and Limitations of AI-Driven Software Testing in Healthcare Applications**

Despite the advancements in AI-powered software testing for healthcare applications, several challenges and limitations must be addressed before widespread adoption. These challenges primarily revolve around data privacy concerns, AI model biases, lack of standardized testing frameworks, high implementation costs, and regulatory complexities. Addressing these limitations is crucial for ensuring the reliability, security, and ethical deployment of AI-driven testing methodologies in healthcare.

* 1. **Data Privacy and Security Concerns**

AI-powered testing often requires access to large volumes of patient data to train machine learning models for defect detection and anomaly identification. However, healthcare data is highly sensitive and protected under strict regulations such as HIPAA (Health Insurance Portability and Accountability Act) in the U.S., GDPR (General Data Protection Regulation) in Europe, and ISO/IEC 27001 (International Information Security Standard). Unauthorized access or data breaches can have severe legal and ethical consequences, making it challenging to use real-world healthcare data for AI model training. Additionally, anonymization techniques used to protect patient data may reduce the effectiveness of AI models, leading to lower accuracy in testing predictions. Ensuring secure, privacy-preserving AI models and adopting techniques such as federated learning (where AI models learn from decentralized data without directly accessing patient records) are potential solutions [27-29].

* 1. **AI Model Bias and Reliability Issues**

One of the major concerns in AI-driven testing is model bias, which can lead to unfair or inaccurate test case prioritization and defect detection. AI models trained on limited or biased datasets may not generalize well across diverse healthcare applications, resulting in false positives or false negatives in test outcomes. For example, an AI-based anomaly detection system trained on a specific hospital's electronic health records (EHR) may not perform well when applied to different healthcare settings with varying patient demographics. This issue is particularly concerning in AI-driven clinical decision support systems (CDSS), where biased test results could lead to incorrect software validations, impacting patient safety. To address this, it is essential to diversify AI training datasets, implement bias-detection algorithms, and adopt explainable AI (XAI) models to enhance transparency in AI-driven testing decisions [30-32].

* 1. **Lack of Standardized AI Testing Frameworks**

Unlike traditional software testing, which follows well-defined standards such as ISTQB (International Software Testing Qualifications Board) and ISO 29119 (Software Testing Standards), AI-driven software testing lacks universal standardization. The absence of standardized AI testing methodologies, performance metrics, and validation benchmarks makes it difficult for healthcare organizations and regulatory bodies to evaluate the effectiveness of AI-powered testing tools. This inconsistency creates challenges in establishing best practices for AI-driven test automation, model verification, and continuous integration (CI/CD) pipelines in healthcare software development. The development of AI-specific software testing frameworks and collaboration between industry leaders, regulatory agencies, and academic institutions will be essential in defining standardized approaches for AI-based testing in healthcare applications [33, 34].

* 1. **High Implementation Costs and Technical Complexity**

Implementing AI-driven software testing requires significant investment in infrastructure, computational resources, and skilled AI professionals. Training and deploying deep learning-based testing models often involve high-performance GPU/TPU computing clusters, which can be expensive for smaller healthcare organizations or startups. Additionally, integrating AI into existing healthcare software testing workflows requires a high level of technical expertise, including knowledge of machine learning algorithms, automated testing frameworks, and data security compliance. The shortage of AI-skilled test engineers further adds to the challenge, making it difficult for organizations to transition from traditional testing methods to AI-powered solutions. Developing cost-effective AI testing frameworks and low-code/no-code AI-driven testing platforms could help mitigate this barrier and make AI-based software validation more accessible [35, 36].

List 2 : The challenges associated with AI-based testing

|  |  |
| --- | --- |
| **Challenge** | **Explanation** |
| **Data Privacy Concerns** | AI-based testing requires access to sensitive patient data, raising privacy issues. |
| **High Initial Setup Cost** | Implementing AI-driven testing requires investment in infrastructure and expertise. |
| **Lack of Standardized Testing Frameworks** | Absence of universal AI-testing standards for healthcare applications. |
| **AI Model Bias** | AI testing tools may inherit biases from training data, affecting test accuracy. |
| **Regulatory Uncertainty** | Compliance with evolving healthcare regulations remains a challenge. |

* 1. **Regulatory** **and Ethical Complexities**

Regulatory compliance is a critical challenge in AI-powered software testing, as healthcare applications must adhere to strict safety, security, and ethical guidelines. AI-driven testing frameworks need to align with FDA (Food and Drug Administration) software validation guidelines, IEC 62304 (Medical Device Software), HIPAA, GDPR, and other healthcare IT regulations [37]. However, current regulatory frameworks were primarily designed for traditional software validation methods and do not fully address the unique challenges posed by AI-driven automation. For instance, AI-powered testing tools may generate self-learning test cases, making it difficult to provide transparent audit trails required by regulatory agencies. Ensuring regulatory compliance in AI testing requires developing AI auditing tools, ethical AI guidelines, and explainability models to provide clear justification for AI-driven test decisions. Additionally, regulatory bodies must evolve their certification and validation processes to accommodate the dynamic nature of AI-powered software testing [38].

* 1. **Limited Real-World Validation and Generalization Issues**

Many AI-based software testing methodologies have been validated in experimental or controlled environments but have limited real-world implementation in healthcare settings. AI models that perform well in simulated testing scenarios may fail when deployed in live healthcare applications due to unforeseen edge cases, software interactions, or variations in healthcare workflows. The lack of large-scale real-world case studies makes it challenging to assess the scalability and adaptability of AI-powered testing tools across diverse healthcare institutions. To overcome this limitation, researchers must focus on cross-validation with real-world clinical trials, longitudinal studies, and large-scale healthcare software implementations to refine AI-driven testing methodologies [39].

* 1. **Ethical Concerns in AI-Driven Testing Decisions**

AI-based testing introduces ethical concerns related to accountability, transparency, and decision-making autonomy. Since AI-driven test automation involves self-learning algorithms, it may generate test cases and validation results that are difficult for human testers to interpret. If an AI-based testing tool incorrectly flags a critical defect in a clinical decision support system, delaying its deployment, or fails to detect a critical software flaw, leading to patient harm, determining accountability becomes complex. There is also the risk of "black-box AI" models making testing decisions without human explainability, making it difficult to trust AI-driven validation processes. Ensuring AI transparency, ethical guidelines, and human-in-the-loop oversight will be essential in addressing these concerns [40].

Fig 4

**5. Future Directions**

As AI-powered software testing continues to evolve, several key research and development areas must be addressed to ensure its successful integration into healthcare applications. Future advancements should focus on standardization, explainability, real-world validation, regulatory compliance, continuous integration with DevOps, cost-effective AI-driven testing frameworks, and ethical AI governance. Addressing these areas will enhance the effectiveness, security, and reliability of AI-driven testing while mitigating existing challenges [41].

1. **Development of Standardized AI Testing Frameworks**

One of the most critical future directions is the establishment of standardized AI testing frameworks for healthcare applications. Unlike traditional software testing, which follows well-defined industry standards such as ISO 29119 (Software Testing Standards) and ISTQB (International Software Testing Qualifications Board), AI-driven testing lacks universally accepted guidelines [42]. The absence of AI-specific testing protocols makes it challenging for healthcare organizations to evaluate the reliability and accuracy of AI-powered test automation tools. To address this, regulatory bodies, research institutions, and industry leaders must collaborate to develop benchmarking frameworks that define:

* AI-driven test case generation and validation methodologies
* Performance metrics for AI-based defect detection and anomaly identification
* Standardized datasets for training and evaluating AI testing tools
* Best practices for integrating AI into automated software testing pipelines

Creating a universal AI software testing standard will help improve interoperability, consistency, and trust in AI-driven healthcare software validation.

**2. Explainable AI (XAI) for Transparent Testing Decisions**

As AI-powered software testing tools become more complex, ensuring transparency and interpretability in test outcomes is essential. Many AI models, particularly deep learning-based ones, operate as black-box systems, making it difficult to understand why certain test cases are prioritized or why defects are flagged. The introduction of Explainable AI (XAI) can enhance trust and usability by providing clear, human-readable explanations of AI-driven test results. Future research should focus on [42]:

* Developing XAI algorithms that justify AI-based test case decisions
* Creating visual dashboards that allow software testers to interpret AI-driven testing insights
* Implementing AI auditing tools that provide traceability for defect detection and validation outcomes
* Enhancing AI transparency to align with regulatory compliance requirements

Explainable AI will not only improve adoption but also help regulatory agencies validate AI-driven testing methodologies in healthcare applications.

**3. Real-World Validation and Large-Scale Implementation**

Most AI-powered software testing methodologies have been validated in controlled experimental environments rather than in real-world clinical settings. However, healthcare applications are highly dynamic, with varying workflows, software architectures, and compliance requirements across different institutions. Future efforts should prioritize large-scale real-world validation by [43]:

* Integrating AI-driven testing into live healthcare IT infrastructures
* Conducting longitudinal studies to assess AI-powered testing performance over time
* Collaborating with hospitals, medical device manufacturers, and telemedicine providers to test AI-driven software validation in practical scenarios
* Developing adaptive AI models that continuously improve testing accuracy based on real-world feedback

By validating AI-driven testing in real-world healthcare environments, organizations can ensure that these tools generalize effectively across different healthcare settings.

**4. Regulatory Compliance and AI-Driven Software Validation**

The integration of AI-driven testing in healthcare applications requires strict regulatory oversight to ensure compliance with safety and security guidelines such as [44]:

* HIPAA (Health Insurance Portability and Accountability Act) – USA
* GDPR (General Data Protection Regulation) – EU
* FDA (Food and Drug Administration) guidelines for software as a medical device (SaMD)
* ISO/IEC 27001 (Information Security Standards) for healthcare IT

Future research should focus on developing AI testing tools that are pre-configured for compliance validation, allowing healthcare organizations to automate regulatory audits. AI-driven compliance assessment frameworks should be designed to:

* Monitor software security vulnerabilities and flag compliance risks in real-time
* Generate automated compliance reports for regulatory approval
* Ensure that AI testing methodologies align with evolving legal and ethical guidelines
* Support dynamic regulatory frameworks that can adapt to changes in AI governance policies

Automating compliance validation through AI will reduce the burden of regulatory audits while ensuring that healthcare applications remain secure and legally compliant.

**5. Integration of AI-Driven Testing with DevOps and Continuous Testing Pipelines**

Future healthcare software development will rely heavily on continuous integration and continuous deployment (CI/CD) pipelines, requiring AI-driven testing to be seamlessly integrated into DevOps workflows. AI-powered test automation tools should be designed to [45]:

* Perform real-time defect detection and code validation in continuous deployment environments
* Dynamically generate test cases based on software updates and new feature releases
* Ensure rapid software iterations without compromising system stability and security
* Predict potential software failures before deployment using AI-driven predictive analytics

By integrating AI-driven testing into CI/CD pipelines, healthcare organizations can achieve faster, more reliable software releases while reducing manual testing overhead.

**6. Cost-Effective AI Testing Frameworks for Small-Scale Healthcare Organizations**

Many small-scale healthcare providers and startups struggle to adopt AI-driven software testing due to high computational costs and resource constraints. Future research should focus on developing cost-effective, scalable AI-powered testing frameworks that [46]:

* Utilize cloud-based AI testing platforms to minimize infrastructure costs
* Leverage federated learning to enable AI training without requiring large, centralized datasets
* Develop lightweight AI models that can run efficiently on limited hardware
* Provide low-code/no-code AI testing solutions to enable non-technical healthcare professionals to implement AI-driven validation

These advancements will help democratize AI-powered testing, making it accessible to a broader range of healthcare providers.

**7. Ethical AI Governance and Responsible AI in Software Testing**

As AI continues to play a critical role in healthcare software validation, ensuring ethical AI governance is crucial. Future research should address [47]:

* Bias mitigation strategies to ensure fairness in AI-powered test automation
* Ethical guidelines for AI-driven software validation in life-critical healthcare applications
* Human-in-the-loop (HITL) models that combine AI automation with human oversight
* Policies to prevent AI misuse in healthcare software testing

By focusing on ethical AI governance, organizations can ensure that AI-powered testing methodologies remain fair, transparent, and aligned with patient safety principles.

Fig 5



**Conclusion**

AI-powered software testing is rapidly transforming the landscape of healthcare applications by improving test efficiency, accuracy, and automation. This systematic review has highlighted the significant benefits of AI-driven software testing, particularly in enhancing test coverage, reducing defect detection time, and ensuring compliance with regulatory standards such as HIPAA, GDPR, and FDA guidelines. By leveraging machine learning, deep learning, natural language processing (NLP), and reinforcement learning, AI-powered testing tools have demonstrated their potential in functional testing, performance testing, security validation, usability testing, and regression testing. Despite these advancements, several challenges remain, including data privacy concerns, AI model biases, lack of standardized testing frameworks, high implementation costs, and regulatory complexities.

One of the key findings of this review is that AI-driven software testing significantly improves test automation and defect prediction, particularly in healthcare applications where system reliability is critical. Studies have shown that AI-powered defect detection achieves accuracy rates exceeding 90% in certain healthcare software validation tasks, significantly reducing software failures post-deployment. Similarly, reinforcement learning techniques have optimized test case generation, leading to better resource allocation and reduced testing costs. Additionally, NLP-based AI models have enhanced the usability testing of AI-driven chatbots, clinical decision support systems (CDSS), and telemedicine applications, ensuring that healthcare AI solutions deliver coherent and contextually accurate outputs.

However, despite its advantages, the adoption of AI-powered software testing in healthcare applications is still hindered by several technical, ethical, and regulatory barriers. One of the primary challenges is data privacy and security concerns, as AI-driven testing often requires access to sensitive patient records and electronic health data. Ensuring compliance with data protection regulations such as HIPAA and GDPR is crucial, and future AI-powered testing frameworks must incorporate privacy-preserving AI techniques, such as federated learning and differential privacy, to mitigate these risks. Another key challenge is AI model bias, where testing algorithms may favor specific datasets, leading to biased defect predictions and unreliable validation outcomes. Addressing bias in AI-driven testing requires diversified training datasets, continuous bias monitoring, and explainable AI (XAI) techniques to ensure transparency and fairness in AI-powered testing decisions.

The review also highlights the lack of standardized AI testing frameworks as a significant limitation. Unlike traditional software testing, which follows well-defined international standards such as ISO 29119 and ISTQB, AI-driven testing lacks universally accepted methodologies for test case generation, defect detection, and validation benchmarks. This inconsistency makes it challenging for regulatory agencies and healthcare organizations to evaluate AI-powered testing tools effectively. Future research should focus on developing standardized AI testing protocols, enabling healthcare institutions to benchmark AI-driven testing performance across different healthcare software applications.

References:

1. *Abadi, M., et al.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308–318. Association for Computing Machinery (2016).*

2. *Ahmed, Z., Mohamed, K., Zeeshan, S., Dong, X.: Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database 2020, baaa010 (2020).*

3. *Alder, S.: Largest healthcare data breaches of 2021. The HIPAA Journal (2021). Accessed 10 Sep 2022.*

4. *Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115 (2020).*

5. *Azevedo, B.F., Rocha, A.M.A., Pereira, A.I.: Hybrid approaches to optimization and machine learning methods: a systematic literature review. Mach. Learn. 1–43 (2024).*

6. *Bell, J.H., Bonawitz, K.A., Gascón, A., Lepoint, T., Raykova, M.: Secure single-server aggregation with (poly) logarithmic overhead. In: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, pp. 1253–1269 (2020).*

7. *Belle, A., Thiagarajan, R., Soroushmehr, S.M., Navidi, F., Beard, D.A., Najarian, K.: Big data analytics in healthcare. BioMed Res. Int. (2015).*

8. *Bertl, M.: Systematic AI support for psychiatry: a framework on how to implement decision support systems. Phd thesis, Tallinn University of Technology, Tallinn, Estonia (November 2023).* [*https://digikogu.taltech.ee/en/Item/a77022c7-b2af-46f9-ae83-62fe04f9bc72*](https://digikogu.taltech.ee/en/Item/a77022c7-b2af-46f9-ae83-62fe04f9bc72)*.*

9. *Bertl, M., Bignoumba, N., Ross, P., Yahia, S.B., Draheim, D.: Evaluation of deep learning-based depression detection using medical claims data. Artif. Intell. Med. (2023).*

10. *Bertl, M., Kankainen, K.J.I., Piho, G., Draheim, D., Ross, P.: Evaluation of data quality in the Estonia national health information system for digital decision support. In: Proceedings of the 3rd International Health Data Workshop. CEUR-WS (2023).*

11. *Bertl, M., Klementi, T., Piho, G., Ross, P., Draheim, D.: How domain engineering can help to raise adoption rates of artificial intelligence in healthcare. In: Delir Haghighi, P., et al. Information Integration and Web Intelligence. iiWAS 2023. LNCS, vol. 14416, pp. 3–12. Springer, Cham (2023).* [*https://doi.org/10.1007/978-3-031-48316-5\_1*](https://doi.org/10.1007/978-3-031-48316-5_1)*.*

12. *Bertl, M., Metsallik, J., Ross, P.: A systematic literature review of AI-based digital decision support systems for post-traumatic stress disorder. Front. Psychiatry 13 (2022).* [*https://doi.org/10.3389/fpsyt.2022.923613*](https://doi.org/10.3389/fpsyt.2022.923613)*.*

13. *Bertl, M., et al.: Future opportunities for systematic AI support in healthcare. In: Steffen, B. (ed.) Bridging the Gap Between AI and Reality - First International Conference, AISoLA 2023, Crete, Greece, 23–28 October 2023, Post-Proceedings (2024). this volume.*

14. *Bertl, M., Ross, P., Draheim, D.: A survey on AI and decision support systems in psychiatry - uncovering a dilemma. Expert Syst. Appl. 202, 117464 (2022).* [*https://doi.org/10.1016/j.eswa.2022.117464*](https://doi.org/10.1016/j.eswa.2022.117464)*.* [*https://www.sciencedirect.com/science/article/pii/S0957417422007965*](https://www.sciencedirect.com/science/article/pii/S0957417422007965)*.*

15. *Bertl, M., Ross, P., Draheim, D.: Systematic AI Support for Decision Making in the Healthcare Sector: Obstacles and Success Factors. Health Policy and Technology (2023).*

16. *Bertl, M., Shahin, M., Ross, P., Draheim, D.: Finding indicator diseases of psychiatric disorders in bigdata using clustered association rule mining. In: Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, pp. 826–833. SAC ’23, Association for Computing Machinery (2023).* [*https://doi.org/10.1145/3555776.3577594*](https://doi.org/10.1145/3555776.3577594)*.*

17. *Ciatto, G., Sabbatini, F., Agiollo, A., Magnini, M., Omicini, A.: Symbolic knowledge extraction and injection with sub-symbolic predictors: a systematic literature review. ACM Comput. Surv. (2024).*

18. *Brandon, C., Singh, A., Margaria, T.: Model driven development for AI-based healthcare systems: a review. In: Steffen, B. (ed.) Bridging the Gap Between AI and Reality - First International Conference, AISoLA 2023, Crete, Greece, 23–28 October 2023, Post-Proceedings (2024). this volume.*

19. *Brandon, C., et al.: Cinco de bio: a low-code platform for domain-specific workflows for biomedical research. BioMedInformatics (to appear) (2024).*

20. *Fennelly, O., Moroney, D., Doyle, M., Eustace-Cook, J., Hughes, M.: Key interoperability factors for patient portals and electronic health records: a scoping review. Int. J. Med. Inform. 105335 (2024).*

21. *Directorate-General for Health and Food Safety: A european health data space: harnessing the power of health data for people, patients and innovation. European Commission (2022).* [*https://health.ec.europa.eu/document/download/17c7065c-c432-445f-9b27-8ccf283581bc\_en?filename=com\_2022-196\_en.pdf*](https://health.ec.europa.eu/document/download/17c7065c-c432-445f-9b27-8ccf283581bc_en?filename=com_2022-196_en.pdf)*. Accessed 28 Jan 2024.*

22. *Floridi, L., et al.: Ai4people–an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind. Mach. 28, 689–707 (2018).*

23. *Fredrikson, M., Jha, S., Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 1322–1333 (2015).*

24. *Gansel, X., Mary, M., van Belkum, A.: Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review. Eur. J. Clin. Microbiol. Infect. Dis. 38(6), 1023—1034 (2019).* [*https://doi.org/10.1007/s10096-019-03501-6*](https://doi.org/10.1007/s10096-019-03501-6)*.*

25. *Gossen, F., Margaria, T., Steffen, B.: Towards explainability in machine learning: the formal methods way. it prof. 22 (4), 8–12 (2020).*

26. *Gossen, F., Margaria, T., Steffen, B.: Formal methods boost experimental performance for explainable ai. it prof. 23 (6), 8–12 (2021).*

27. *Kavitha, M., Srinivas, P.V.V.S., Kalyampudi, P., F, C.S., Srinivasulu, S.: Machine learning techniques for anomaly detection in smart healthcare. In: 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1350–1356 (2021).* [*https://doi.org/10.1109/ICIRCA51532.2021.9544795*](https://doi.org/10.1109/ICIRCA51532.2021.9544795)*.*

28. *Khanna, N.N., et al.: Economics of artificial intelligence in healthcare: Diagnosis vs. treatment. Healthcare 10(12) (2022).* [*https://doi.org/10.3390/healthcare10122493*](https://doi.org/10.3390/healthcare10122493)*.* [*https://www.mdpi.com/2227-9032/10/12/2493*](https://www.mdpi.com/2227-9032/10/12/2493)*.*

29. *Klementi, T., Kankainen, K.J.I., Piho, G., Ross, P.: Prospective research topics towards preserving electronic health records in decentralised content-addressable storage networks. In: HEDA@ Petri Nets, p. 14 (2022).*

30. *Klementi, T., Piho, G.: Method and system for managing data using decentralized content-addressable storage networks. European patent Office, priority number EP24166173.5 (March 2024). submitted pattent.*

31. *Klementi, T., Piho, G., Ross, P.: A reference architecture for resolving health data accessibility, comprehensiveness, and ownership dilemmas through preserving semantically interoperable personal health records in decentralized content-addressable storage networks. (2024), will be published soon.*

32. Asan, O., A.E. Bayrak, and A. Choudhury, *Artificial intelligence and human trust in healthcare: focus on clinicians.* J. Med. Internet Res., 2020. **22**.

33. Asch, M., *Big data and extreme-scale computing: pathways to convergence-toward a shaping strategy for a future software and data ecosystem for scientific inquiry.* Int. J. High Perform. Comput. Appl., 2018. **32**.

34. Bertl, M., *News analysis for the detection of cyber security issues in digital healthcare: a text mining approach to uncover actors, attack methods and technologies for cyber defense.* Young Inf. Sci., 2019. **4**.

35. Bertl, M., P. Ross, and D. Draheim, *Predicting psychiatric diseases using AutoAI: a performance analysis based on health insurance billing data*, in *Database and Expert Systems Applications*, C. Strauss, et al., Editors. 2021, Springer: Cham.

36. Bhatt, P., et al., *Emerging artificial intelligence-empowered mhealth: scoping review.* JMIR Mhealth Uhealth, 2022. **10**.

37. Gerke, S., et al., *The need for a system view to regulate artificial intelligence/machine learning-based software as medical device.* NPJ Digit. Med., 2020. **3**.

38. Gubala, V., et al., *Point of care diagnostics: status and future.* Anal. Chem., 2012. **84**.

39. Klar, E., *Medical Device Regulation als aktuelle Herausforderung für die rechtssichere Einführung neuer Technologien.* Chirurg, 2018. **89**.

40. Hidayat, M., S.Y. Defitri, and H. Hilman, *The impact of artificial intelligence (AI) on financial management.* Manag. Stud. Bus. J. (PRODUCTIVITY), 2024. **1**.

41. Wang, Q., et al., *Integrating digital technologies and public health to fight COVID-19 pandemic: key technologies, applications, challenges and outlook of digital healthcare.* Int. J. Environ. Res. Public Health, 2021. **18**.

42. Strickland, E., *Ibm watson, heal thyself: how ibm overpromised and underdelivered on ai health care.* IEEE Spectr., 2019. **56**.

43. *Fujii, Y.R.: The MicroRNA 2000 Transformer: Quantum Computing and Artificial Intelligence for Health. Springer Nature (2023).* [*https://doi.org/10.1007/978-981-99-3165-1*](https://doi.org/10.1007/978-981-99-3165-1)*.*

44. *Gupta, R., et al.: VAHAK: a blockchain-based outdoor delivery scheme using UAV for Healthcare 4.0 services. In: IEEE INFOCOM 2020 – IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 255–260. IEEE (2020).* [*https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162738*](https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162738)*.*

45. *Gupta, R., Shukla, A., Tanwar, S.: Aayush: a smart contract-based telesurgery system for healthcare 4.0. In: 2020 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6. IEEE (2020).* [*https://doi.org/10.1109/ICCWorkshops49005.2020.9145044*](https://doi.org/10.1109/ICCWorkshops49005.2020.9145044)*.*

46. *International Electrotechnical Commission (IEC): IEC 60601-1-2:2014 - IEC-Normen - VDE VERLAG.* [*https://www.vde-verlag.de/iec-normen/220613/iec-60601-1-2-2014.html*](https://www.vde-verlag.de/iec-normen/220613/iec-60601-1-2-2014.html)*. Accessed 07 Dec 2023.*

47. *International Electrotechnical Commission (IEC): IEC62304:2006/AMD1:2015 amendment 1-medical device software-software life cycle processes (2015).* [*https://webstore.iec.ch/publication/22790*](https://webstore.iec.ch/publication/22790)*.*