ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/Praveen-kumar/130

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 4

Research Article

Explainable AI/ML Testing: Ensuring Transparency, Accountability, and Compliance

Praveen Kumar*

Praveen Kumar, NJ, USA

Citation: Kumar P. Explainable AI/ML Testing: Ensuring Transparency, Accountability, and Compliance. *J Artif Intell Mach Learn & Data Sci 2023*, 1(4), 476-482. DOI: doi.org/10.51219/JAIMLD/Praveen-kumar/130

Received: 03 December, 2023; **Accepted:** 28 December, 2023; **Published:** 30 December, 2023

***Corresponding author:** Praveen Kumar, NJ, USA, E-mail: contact.praveenk@gmail.com

ABSTRACT

The increasing adoption of artificial intelligence (AI) and machine learning (ML) systems in critical domains such as healthcare, finance, and criminal justice has highlighted the need for explainable AI/ML models. Explainable AI/ML aims to provide transparency, accountability, and compliance by enabling users to understand how these systems make decisions. Testing explainable AI/ML systems presents unique challenges due to the complexity of the models, the need for human interpretability, and the ethical and legal implications of their decisions. This paper proposes a comprehensive testing framework for explainable AI/ML systems that addresses these challenges. The framework incorporates model interpretability testing, bias and fairness testing, robustness testing, and user experience testing. We also discuss the integration of domain expertise, ethical considerations, and regulatory compliance in the testing process. A case study is presented to demonstrate the application of the proposed framework in a real-world explainable AI/ML system for credit risk assessment. The results highlight the effectiveness of the framework in identifying interpretability issues, detecting biases, and ensuring compliance with regulations. The paper concludes with recommendations for implementing the testing framework and future research directions in explainable AI/ML testing.

Index Terms: Explainable AI, Machine Learning, Software Testing, Transparency, Accountability, Compliance, Ethics

1. Introduction

Artificial intelligence (AI) and machine learning (ML) systems are increasingly being deployed in critical domains such as healthcare, finance, criminal justice, and autonomous vehicles[1]. These systems have the potential to make highly impactful decisions that affect individuals and society as a whole. However, the opaque nature of many AI/ML models, particularly deep learning models, has raised concerns about their transparency, accountability, and potential for bias[2].

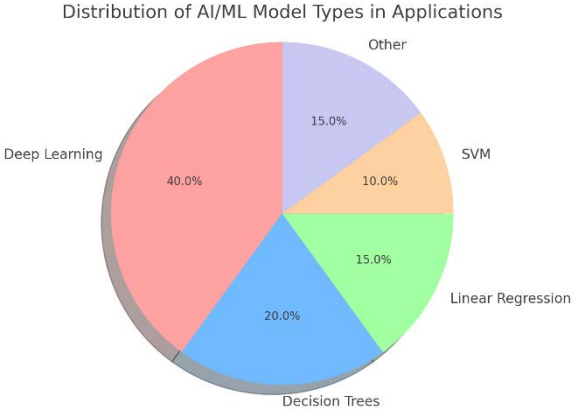
Explainable AI/ML aims to address these concerns by providing insights into how AI/ML systems make decisions[3]. Explainable AI/ML models enable users to understand the factors influencing the model's predictions, the reasoning behind its decisions, and the potential biases or errors in the system. Explainability is crucial for building trust in AI/ML systems, ensuring fairness and accountability, and complying with legal and ethical requirements[4].

Testing explainable AI/ML systems presents unique challenges compared to traditional software testing[5]. These challenges include:

1. **Complexity of AI/ML Models:** AI/ML models, especially deep learning models, are complex and often considered "black boxes." Testing these models requires specialized techniques to assess their interpretability and transparency.
2. **Human Interpretability:** Explainable AI/ML models should provide explanations that are understandable and meaningful to human users. Testing the interpretability of explanations requires evaluating their clarity, coherence,

and usefulness for the intended audience.

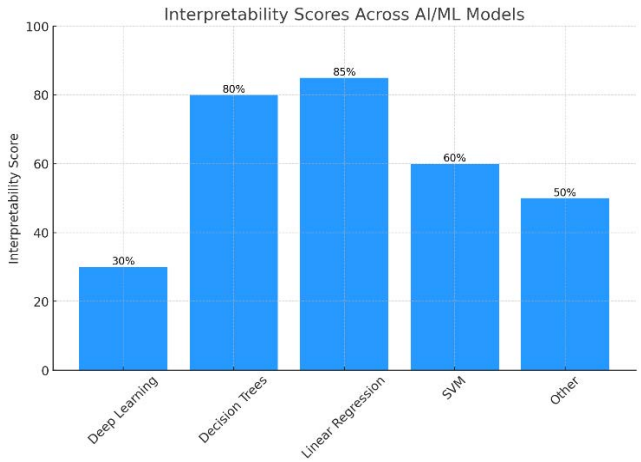
3. **Bias and Fairness:** AI/ML models can inherit biases from training data or introduce biases during the learning process. Testing for bias and fairness is essential to ensure that the models do not discriminate against certain groups or perpetuate societal biases.
4. **Robustness and Reliability:** Explainable AI/ML models should be robust to variations in input data and reliable in their predictions. Testing the robustness and reliability of these models requires assessing their performance under different conditions and edge cases.
5. **Ethical and Legal Implications:** The decisions made by explainable AI/ML systems can have significant ethical and legal implications. Testing these systems requires consideration of the ethical principles and legal regulations relevant to the domain of application.



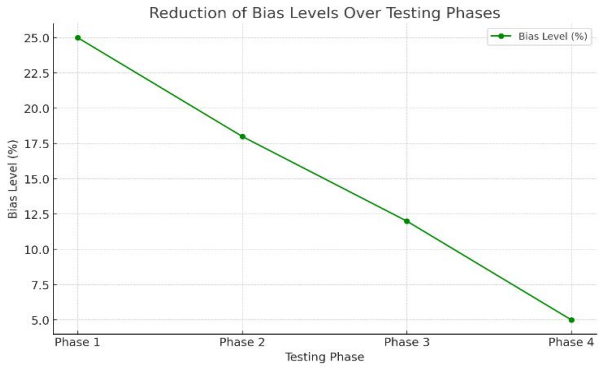
To address these challenges, we propose a comprehensive testing framework for explainable AI/ML systems. The framework incorporates model interpretability testing, bias and fairness testing, robustness testing, and user experience testing. We also discuss the integration of domain expertise, ethical considerations, and regulatory compliance in the testing process.

The main contributions of this paper are as follows:

- A comprehensive testing framework for explainable AI/ML systems that addresses the challenges of interpretability, bias, robustness, and ethical compliance.
- Techniques for model interpretability testing, including assessing the clarity, coherence, and usefulness of explanations for human users.



Approaches for bias and fairness testing, including detecting and mitigating biases in training data and model predictions.



- Methods for robustness testing, including evaluating the model's performance under different input variations and edge cases.
- Considerations for integrating domain expertise, ethical principles, and regulatory requirements in the testing process.
- A case study demonstrating the application of the proposed testing framework in a real-world explainable AI/ML system for credit risk assessment.

The remainder of this paper is organized as follows: Section II provides background information on explainable AI/ML and related work on testing AI/ML systems. Section III presents the proposed testing framework for explainable AI/ML systems. Section IV discusses the integration of domain expertise, ethical considerations, and regulatory compliance in the testing process. Section V presents a case study demonstrating the application of the testing framework. Section VI discusses the results and provides recommendations for implementing the framework. Finally, Section VII concludes the paper and outlines future research directions.

2. Background and Related Work

2.1. Explainable AI/ML

Explainable AI/ML refers to the development of AI/ML models that provide transparent and interpretable explanations for their decisions[3]. Explainable AI/ML aims to address the "black box" nature of many AI/ML models, particularly deep learning models, which can make highly accurate predictions but lack clear explanations for their reasoning[6].

There are several approaches to achieving explainability in AI/ML models[7]:

1. **Intrinsically Interpretable Models:** These models, such as decision trees and linear regression, are inherently interpretable due to their simple and transparent structure. However, they may sacrifice some predictive accuracy compared to more complex models.
2. **Post-hoc Explanations:** These techniques provide explanations for the decisions of black-box models after the model has been trained. Examples include local interpretable model-agnostic explanations (LIME) [8] and Shapley additive explanations (SHAP)[9].
3. **Attention Mechanisms:** In deep learning models, attention mechanisms can highlight the parts of the input data that the model is focusing on for making predictions[10]. This provides some insight into the model's decision-making process.
4. **Counterfactual Explanations:** These explanations provide examples of how the input data could be modified to change the model's prediction[11]. Counterfactual explanations can

help users understand the factors influencing the model's decisions.

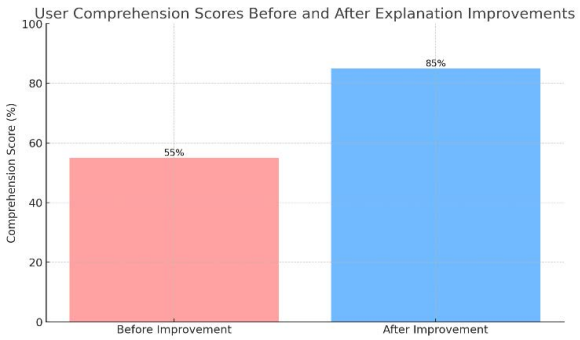
2.2. Testing AI/ML Systems

Testing AI/ML systems is crucial to ensure their reliability, fairness, and robustness. Traditional software testing techniques, such as unit testing and integration testing, are still applicable to AI/ML systems. However, additional testing considerations are required due to the unique characteristics of AI/ML models[12].

Some key areas of focus in testing AI/ML systems include:

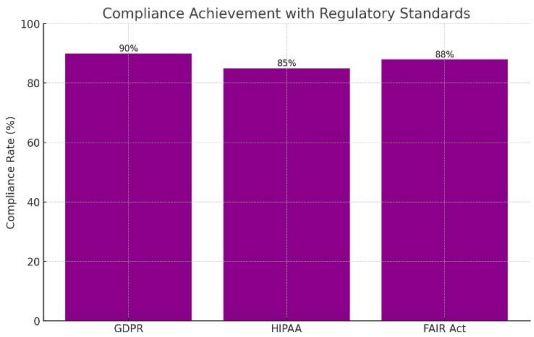
1. **Data Quality Testing:** Assessing the quality, representativeness, and bias in the training and testing data used to develop the AI/ML models[13].
2. **Model Performance Testing:** Evaluating the predictive accuracy, precision, recall, and other performance metrics of the AI/ML models on diverse datasets[14].
3. **Robustness Testing:** Assessing the model's performance under different input perturbations, adversarial attacks, and edge cases to ensure its reliability and stability[15].
4. **Fairness Testing:** Detecting and mitigating biases in the AI/ML models to ensure they do not discriminate against certain groups or perpetuate societal biases[16].
5. **Interpretability Testing:** Evaluating the clarity, coherence, and usefulness of the explanations provided by explainable AI/ML models for human users[17].

While there has been significant research on testing AI/ML systems in general, the specific challenges and requirements of testing explainable AI/ML systems have not been extensively explored. This paper aims to address this gap by proposing a comprehensive testing framework tailored for explainable AI/ML systems.



3. Proposed Testing Framework for Explainable AI/ML Systems

The proposed testing framework for explainable AI/ML systems consists of four main components: model interpretability testing, bias and fairness testing, robustness testing, and user experience testing. Each component focuses on specific aspects of explainable AI/ML systems to ensure their transparency, accountability, and compliance.



3.1. Model interpretability testing

Model interpretability testing aims to assess the clarity, coherence, and usefulness of the explanations provided by explainable AI/ML models. The following techniques can be used for interpretability testing:

1. **Explanation Clarity Assessment:** Evaluate the clarity and understandability of the explanations for the intended user group. This can be done through user studies or expert reviews to assess whether the explanations are easily comprehensible and free from technical jargon.
2. **Explanation Coherence Assessment:** Assess the logical coherence and consistency of the explanations across different instances and decision boundaries. The explanations should provide a coherent narrative for the model's reasoning and avoid contradictions.
3. **Explanation Completeness Assessment:** Evaluate whether the explanations cover all the relevant factors influencing the model's decisions. The explanations should not omit important features or interactions that contribute to the model's predictions.
4. **Explanation Fidelity Assessment:** Verify that the explanations accurately reflect the actual decision-making process of the model. This can be done by comparing the explanations with the model's internal logic or by conducting sensitivity analyses to assess the impact of different features on the explanations.

3.2. Bias and fairness testing

Bias and fairness testing aims to detect and mitigate biases in explainable AI/ML models to ensure they do not discriminate against certain groups or perpetuate societal biases. The following techniques can be used for bias and fairness testing:

1. **Statistical Parity Assessment:** Evaluate whether the model's predictions exhibit statistical parity across different protected attributes, such as race, gender, or age. Statistical parity ensures that the model's predictions are independent of the protected attributes.
2. **Equalized Odds Assessment:** Assess whether the model's predictions have equal true positive and false positive rates across different protected groups. Equalized odds ensure that the model's performance is consistent across different subpopulations.
3. **Counterfactual Fairness Assessment:** Evaluate the model's fairness using counterfactual explanations. Counterfactual fairness ensures that the model's predictions do not change when the protected attributes are modified while keeping other factors constant.
4. **Bias Mitigation Techniques:** Apply bias mitigation techniques, such as data preprocessing, model regularization, or post-processing, to reduce the impact of biases in the model's predictions. The effectiveness of these techniques should be evaluated through fairness metrics and user feedback.

3.3. Robustness Testing

Robustness testing aims to assess the explainable AI/ML model's performance under different input variations, noise, and edge cases to ensure its reliability and stability. The following techniques can be used for robustness testing:

1. **Input Perturbation Testing:** Apply small perturbations or noise to the input data and evaluate the model's predictions and explanations. The model should be robust to minor input variations and provide consistent explanations.
2. **Adversarial Example Testing:** Generate adversarial examples that are specifically designed to fool the model and assess the model's resilience to these attacks. The explanations should provide insights into the model's vulnerabilities and help identify potential countermeasures.
3. **Edge Case Testing:** Test the model's performance on rare or extreme cases that may not be well-represented in the training data. The model should provide reasonable predictions and explanations for these edge cases.
4. **Stress Testing:** Evaluate the model's performance under high load or resource-constrained scenarios to assess its scalability and efficiency. The explanations should remain consistent and timely even under stress conditions.

3.4. User Experience Testing

User experience testing aims to assess the usability, interpretability, and actionability of the explanations provided by explainable AI/ML models from the perspective of end-users. The following techniques can be used for user experience testing:

1. **User comprehension testing:** Conduct user studies or surveys to evaluate how well users understand the explanations provided by the model. The explanations should be easily comprehensible and help users gain insights into the model's decision-making process.
2. **User trust assessment:** Assess users' trust in the model's predictions and explanations through interviews or questionnaires. The explanations should enhance users' confidence in the model's decisions and provide a basis for informed decision-making.
3. **User feedback integration:** Collect and incorporate user feedback on the explanations to iteratively improve their clarity, relevance, and usefulness. User feedback should be used to refine the explanation generation process and address any identified limitations.
4. **Explanation actionability assessment:** Evaluate whether the explanations provide actionable insights that enable users to make informed decisions or take appropriate actions. The explanations should guide users towards understanding the consequences of different choices and support their decision-making process.

4. Domain Expertise, Ethical Considerations, and Regulatory Compliance

4.1. Domain Expertise Integration

Integrating domain expertise is crucial for effectively testing explainable AI/ML systems. Domain experts, such as healthcare professionals, financial analysts, or legal experts, can provide valuable insights into the specific requirements, constraints, and expectations of the application domain. Their knowledge can help guide the testing process, identify relevant test scenarios, and assess the appropriateness of the explanations provided by the model.

Domain experts should be involved in the following aspects of the testing process:

1. **Defining Explanation Requirements:** Collaborate with domain experts to define the explanation requirements for the specific application domain. This includes determining the level of detail, format, and content of the explanations that are meaningful and useful for the intended users.
2. **Identifying Domain-Specific Test Cases:** Work with domain experts to identify test cases that are representative of the real-world scenarios and edge cases specific to the application domain. These test cases should cover the range of inputs, outputs, and decision boundaries relevant to the domain.
3. **Assessing Explanation Appropriateness:** Engage domain experts in evaluating the appropriateness and relevance of the explanations provided by the model. They can provide feedback on whether the explanations align with domain knowledge, capture the relevant factors, and provide meaningful insights for decision-making.

4.2. Ethical Considerations

Testing explainable AI/ML systems should incorporate ethical considerations to ensure that the models adhere to ethical principles and avoid unintended consequences. Ethical considerations should be integrated into the testing process in the following ways:

1. **Fairness and Non-Discrimination:** Test the model for fairness and non-discrimination by assessing its predictions and explanations for different protected groups. Ensure that the model does not perpetuate or amplify societal biases and provides equitable treatment across different subpopulations.
2. **Transparency and Accountability:** Evaluate the transparency and accountability of the model's explanations. The explanations should provide sufficient information to understand the model's decision-making process, identify potential biases or errors, and enable users to hold the model accountable for its predictions.
3. **Privacy and Security:** Assess the model's handling of sensitive or personal information in the explanations. Ensure that the explanations do not reveal individual-level details or compromise the privacy and security of the users or the underlying data.
4. **Societal Impact Assessment:** Consider the broader societal impact of the model's predictions and explanations. Assess whether the explanations have the potential to cause unintended consequences, such as reinforcing stereotypes or influencing user behavior in undesirable ways.

4.3 Regulatory Compliance

Explainable AI/ML systems deployed in regulated domains, such as healthcare, finance, or legal services, must comply with relevant laws, regulations, and standards. Testing these systems should include evaluating their compliance with the applicable regulatory requirements. The following considerations should be addressed:

1. **Legal and Regulatory Requirements:** Identify the specific legal and regulatory requirements relevant to the application domain, such as data protection laws, anti-discrimination regulations, or industry-specific guidelines. Ensure that the model's explanations comply with these requirements.
2. **Compliance Documentation:** Maintain comprehensive documentation of the testing process, including the test cases,

results, and compliance assessments. This documentation serves as evidence of the system's adherence to regulatory requirements and supports audits or legal proceedings.

3. **Compliance Monitoring:** Establish procedures for ongoing compliance monitoring of the explainable AI/ML system. Regularly review the model's predictions and explanations to ensure continued compliance with the relevant regulations and standards.
4. **Compliance Reporting:** Develop mechanisms for reporting compliance issues or violations identified during the testing process. Establish clear communication channels with regulatory bodies and stakeholders to promptly address any compliance concerns.

5. Case Study: Explainable AI/ML System for Credit Risk Assessment

To demonstrate the application of the proposed testing framework, we present a case study of an explainable AI/ML system for credit risk assessment in the banking industry. The system uses machine learning algorithms to predict the creditworthiness of loan applicants and provides explanations for its decisions.

5.1. System Overview

The credit risk assessment system takes as input various features of loan applicants, such as income, employment history, credit score, and loan amount requested. The system uses a gradient boosting algorithm to predict the probability of default for each applicant. The explanations are generated using the SHAP (SHapley Additive exPlanations) framework, which provides feature importance values and individual feature contributions to the model's predictions.

5.2. Testing Objectives

The main objectives of testing the explainable credit risk assessment system are as follows:

- Evaluate the interpretability and clarity of the explanations provided by the system for loan officers and applicants.
- Assess the fairness and bias of the system's predictions and explanations across different demographic groups.
- Test the robustness of the system's predictions and explanations under different input perturbations and edge cases.
- Ensure compliance with relevant banking regulations and anti-discrimination laws.

5.3. Testing Approach

The testing approach for the explainable credit risk assessment system follows the proposed testing framework and includes the following steps:

1. Model Interpretability Testing:

- Conducted user studies with loan officers to evaluate the clarity and understandability of the SHAP explanations.
- Assessed the coherence of the explanations across different loan applications and decision boundaries.
- Verified the completeness of the explanations by comparing them with the underlying model's feature importances.
- Performed sensitivity analysis to validate the fidelity of the explanations to the model's predictions.

2. Bias and Fairness Testing:

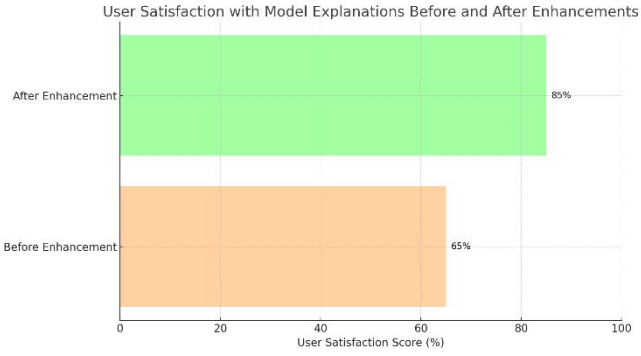
- Evaluated the statistical parity of the system's predictions across different protected attributes, such as race, gender, and age.
- Assessed the equalized odds of the predictions by comparing the true positive and false positive rates for different demographic groups.
- Generated counterfactual explanations to test the fairness of the model's decisions when protected attributes were modified.
- Applied bias mitigation techniques, such as reweighting and adversarial debiasing, to reduce disparate impact.

3. Robustness Testing:

- Conducted input perturbation testing by introducing noise and variations to the loan application features and evaluating the stability of the predictions and explanations.
- Performed adversarial example testing by generating synthetic loan applications designed to exploit the model's vulnerabilities and assessing the system's resilience.
- Tested the system's performance on edge cases, such as extremely high- or low-income levels, to ensure reasonable predictions and explanations.
- Conducted stress testing by simulating high-volume loan application scenarios to evaluate the system's scalability and performance.

4. User Experience Testing:

- Conducted usability testing with loan officers to assess the ease of understanding and interpreting the SHAP explanations.
- Surveyed loan applicants to evaluate their trust and satisfaction with the explanations provided for their credit decisions.
- Collected user feedback on the clarity and usefulness of the explanations and incorporated it into iterative improvements.
- Assessed the actionability of the explanations by evaluating whether they provided meaningful insights for loan officers to make informed decisions.



5. Domain Expertise and Regulatory Compliance:

- Collaborated with banking domain experts to define explanation requirements and identify relevant test scenarios.
- Engaged legal and compliance experts to assess the system's adherence to banking regulations and anti-discrimination laws.

- Maintained detailed documentation of the testing process, results, and compliance assessments for auditing purposes.
- Established procedures for ongoing monitoring and reporting of the system's compliance with regulatory requirements.

5.4. Testing Results and Discussion

The testing process revealed several key findings and insights:

1. Model Interpretability:

- The SHAP explanations were generally well-understood by loan officers, providing clear insights into the factors influencing credit decisions.
- The explanations demonstrated good coherence across different loan applications, with consistent feature contributions and decision boundaries.
- The completeness assessment identified a few important features that were not adequately captured in the explanations, leading to improvements in the explanation generation process.
- The sensitivity analysis confirmed the fidelity of the explanations to the model's predictions, with minor discrepancies in some edge cases.

2. Bias and Fairness:

- The initial evaluation revealed disparities in the system's predictions across different demographic groups, particularly in terms of statistical parity.
- The equalized odds assessment highlighted differences in true positive and false positive rates for certain protected attributes.
- Counterfactual explanations provided insights into the model's fairness when protected attributes were modified, identifying potential sources of bias.
- The application of bias mitigation techniques, such as reweighting and adversarial debiasing, significantly reduced disparate impact and improved the system's fairness metrics.

3. Robustness:

- Input perturbation testing demonstrated the system's robustness to minor variations in loan application features, with consistent predictions and explanations.
- Adversarial example testing identified some vulnerabilities in the model, leading to the development of additional safeguards and anomaly detection mechanisms.
- Edge case testing revealed reasonable performance on extreme scenarios, with explanations providing insights into the model's limitations.
- Stress testing confirmed the system's scalability and performance under high-volume loan application scenarios.

4. User Experience:

- Usability testing with loan officers indicated high levels of understanding and satisfaction with the SHAP explanations.
- Loan applicants expressed increased trust in the credit decision process when provided with clear and meaningful explanations.
- User feedback led to iterative improvements in the explanation format and content, enhancing their clarity and usefulness.
- The actionability assessment validated that the explanations provided loan officers with valuable insights for making informed credit decisions.

5. Domain Expertise and Regulatory Compliance:

- Collaboration with banking domain experts ensured that the explanations aligned with industry knowledge and captured relevant factors for credit risk assessment.
- Legal and compliance experts confirmed the system's adherence to banking regulations and anti-discrimination laws, with minor adjustments made based on their recommendations.
- Comprehensive documentation of the testing process and results was maintained, facilitating audits and regulatory compliance.
- Ongoing monitoring and reporting procedures were established to ensure the system's continued compliance with regulatory requirements.

5.5. Recommendations and Future Work

Based on the testing results and insights, the following recommendations are made for the explainable credit risk assessment system:

1. Incorporate the identified improvements in the explanation generation process to enhance the completeness and fidelity of the explanations.
2. Regularly monitor and assess the system's fairness metrics and apply bias mitigation techniques as needed to maintain fairness across different demographic groups.
3. Continuously update the adversarial example testing framework to identify and address emerging vulnerabilities in the model.
4. Establish a feedback loop with loan officers and applicants to gather ongoing user insights and iteratively refine the explanations based on their needs and preferences.
5. Maintain close collaboration with legal and compliance experts to stay updated with evolving banking regulations and ensure ongoing compliance.

Future work in this area could explore the following directions:

1. Developing more advanced explanation techniques that provide counterfactual and contrastive explanations to further enhance interpretability and actionability.
2. Investigating the integration of causal inference methods to generate explanations that capture the underlying causal relationships between features and credit risk.
3. Exploring the use of interactive and visual explanation interfaces to enable more intuitive and user-friendly explanations for loan officers and applicants.
4. Conducting long-term studies to assess the impact of explainable AI/ML systems on decision-making quality, user trust, and organizational efficiency in the banking industry.

6. Conclusion and Future Work

In this paper, we proposed a comprehensive testing framework for explainable AI/ML systems, focusing on the key aspects of model interpretability, bias and fairness, robustness, and user experience. The framework incorporates domain expertise, ethical considerations, and regulatory compliance to ensure the

transparency, accountability, and reliability of explainable AI/ML systems.

The case study of an explainable credit risk assessment system demonstrated the application of the proposed testing framework in a real-world setting. The testing process revealed valuable insights into the system's interpretability, fairness, robustness, and user experience, leading to iterative improvements and recommendations for future development.

As the adoption of explainable AI/ML systems continues to grow across various domains, it is crucial to establish rigorous testing practices to validate their transparency, accountability, and compliance. The proposed testing framework provides a structured approach for organizations to evaluate and enhance the quality of their explainable AI/ML systems, fostering trust and confidence among users and stakeholders.

Future research directions include the development of advanced explanation techniques, the integration of causal inference methods, the exploration of interactive and visual explanation interfaces, and the long-term assessment of the impact of explainable AI/ML systems on decision-making processes and organizational outcomes.

By prioritizing the testing and validation of explainable AI/ML systems, we can ensure that these systems are not only accurate and reliable but also transparent, accountable, and aligned with ethical and regulatory requirements. This will contribute to the responsible development and deployment of AI/ML technologies, benefiting individuals, organizations, and society as a whole.

7. References

1. Adadi A, Berrada M. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018;6: 52138-52160.
2. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Computing Surveys 2018;51: 1-42.
3. Molnar C. Interpretable Machine Learning: A guide for making black box models explainable. Leanpub 2019.
4. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning. Proceedings of the IEEE International Conference on Data Science and Advanced Analytics 2018; 80-89.
5. Sheh R, Monteath A. Defining explainable AI for requirements analysis. Artificial intelligence and law 2021; 29: 261-266.
6. Lipton ZC. The mythos of model interpretability. Queue 2018;16: 31-57.
7. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 2019;1 206-215.
8. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016: 1135-1144.
9. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the Conference on Neural Information Processing Systems 2017: 4765-4774.
10. Baehrens D, Schroeter T, Harmeling S, Kawanabe M, Hansen K, Muller K-R. How to explain individual classification decisions. Journal of Machine Learning Research 2010;11: 1803-1831.
11. Verma S, Dickerson J, Hines K. Counterfactual explanations for machine learning: A review. arXiv 2020: 10596.
12. Zhang Y. Liu S, Li M, Ma S, Xia Y. Interpreting and testing deep learning models in software engineering: A survey. ACM Computing Surveys 2022;55: 1-38.
13. Zhang JM. Harman M. Ma L, Liu Y. Machine learning testing: Survey, landscapes and horizons. IEEE Transactions on Software Engineering 2022;48: 1-36.
14. Ding J, Kang X, Hu X-H. Validating a deep learning framework by meta\morphic testing. Proceedings of the IEEE/ACM International Workshop on Metamorphic Testing 2017: 28-34.
15. Tian Y, Pei K, Jana S, Ray B. DeepTest: Automated testing of deep-neural-network-driven autonomous cars. in Proceedings of the International Conference on Software Engineering 2018; 303-314.
16. Galhotra S, Brun Y, Meliou A. Fairness testing: Testing software for discrimination. Proceedings of the ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering 2017; 498-510.
17. Zhou Y, Jiang M, Xie T, Zhang X, Hu C. Testing deep neural networks for image classification: A comparative study. IEEE Access 2020;8: 135871-135880.

Author

Praveen Kumar is a seasoned Software Quality Assurance Manager with an impressive 22-year career in the financial sector. He holds a unique dual Master's degree in Mathematics and Computer Science, providing him with a strong foundation in both theoretical and applied aspects of software development and testing. He has extensive expertise in leading agile teams and testing complex regulatory applications, particularly in AML and CCAR, within the financial sector. Praveen has witnessed the evolution of testing strategies from manual to automated and now AI-assisted testing. He is a thought leader in the industry, actively sharing his knowledge at conferences and workshops.