Unsanctioned Intelligence: Mitigating the Risks of 'Shadow AI' and Proprietary Data Leaks in the Remote Workforce

Adel Saad Assiri
Department of Business Informatics, College of Business
King Khaled University, Abha, 61421, Saudi Arabia

***Abstract*—The high implementation of remote working models has only increased the pace of using artificial intelligence tools by employees to improve their productivity, automate routine tasks, and aid in decision-making. Non-committal application of AI tools, or Shadow AI, has, however, become a major organizational risk, especially when the processed information is sensitive, proprietary, or controlled and not handled in line with the existing organizational policies. The current mitigation is based on fairly flat policies, manual checks, or limited monitoring systems, which cannot provide real-time awareness, dynamic risk response, or gated controls within the dynamic remote workforce. To minimize the leaking of proprietary data, this study suggests a single framework of Shadow AI risk mitigation to detect, evaluate, and control unsanctioned AI utilisation. The architecture incorporates automated scoring of risks, policy-sensitive enforcement, risk uncertainty estimation, and governance-driven adaptation as a way of balancing security, usability, and operational efficiency. Experimental results based on synthetic enterprise data (20,000 activity records) and real-world organizational analytics (7,500 interaction records) show that it can detect jobs with an accuracy of 91.6, reduce high-risk Shadow AI interactions by a factor of 62.6 and reduce policy violations by half, using less than 16% computation overhead. The suggested solution offers a scaling- and governance-based solution to securing the use of AI solutions in distributed working scenarios.**

Keywords*—Shadow AI; remote workforce security; data leakage prevention; AI governance; trustworthy AI; enterprise risk management; policy compliance; uncertainty-aware systems

I. Introduction

Remote and hybrid working has increased the adoption of artificial intelligence tools by employees to enhance their productivity and decision support. Although these tools are beneficial in terms of operations, their use without authorization, also known as Shadow AI, has proven to be a major organizational threat in cases where AI systems are implemented without established enterprise governance and oversight systems [1], [2].

Shadow AI has the potential to leak proprietary, confidential, or regulated information onto external platforms, which is a critical threat. Remote working conditions strengthen this threat due to reduced control over centralization, heightened use of personal devices, and diminished identification of authorized and foreign services [3], [4]. Workers can easily provide third-party artificial intelligence tools with confidential data without realizing that data will be stored, trained, or used to comply with regulations, without being aware that the data will be leaked, which will cause intellectual property theft and violations of regulatory policies [5].

Current enterprise responses are isolated and are mostly in reactive mode. The policy-based controls rely on voluntary compliance and are not enforceable, whereas the technical controls, like data leakage prevention and network monitoring, provide a limited insight into AI-specific interactions [6], [7]. Besides, existing AI governance systems are mostly centred on approved systems, which have no control over Shadow AI practices [8]. These restrictions are increased by the fact that there is no uncertainty-based risk assessment since most systems are based on set rules and methods that lack contextual variability and trust in enforcement action [9].

In an attempt to curb these issues, the current study suggests a combined system of Shadow AI threats within the setting of remote labour. This design integrates automated detection, modelling of the risk that humans adapt to, policy-oriented enforcement, and uncertainty to allow real-time, transparent, and governance-orientated control of unsanctioned AI use. The experimental assessment of the proposed framework supports the idea of the work and proves that the risk can be reduced effectively without impacting the Fig 1. shows how unsanctioned AI use in a remote workforce is identified, assessed for risk, governed through policy enforcement, and guided toward secure and compliant AI usage.

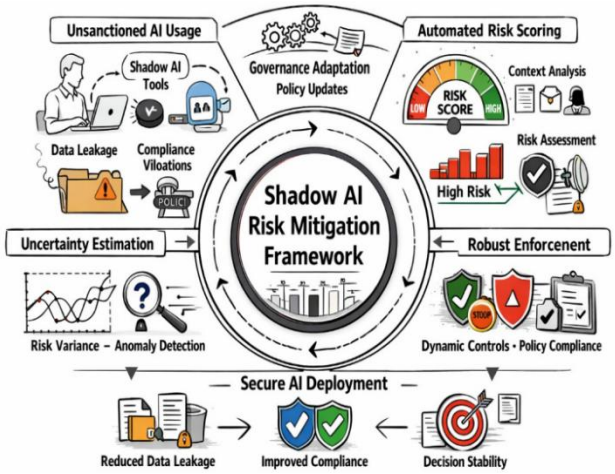

Fig. 1. Graphical overview of the proposed Shadow AI-risk mitigation framework for remote workforces.

The rest of this study is structured in the following way. Section II provides a literature review on the topic of Shadow AI, remote workforce security, and currently used governance and mitigation strategies. Section III entails the problem statement and sets the objectives of the research. Section IV captures the proposed risk mitigation framework and methodology. Section V presents the protocol of the experiment and evaluation. Section VI provides the discussion of the results and performance analysis. Conclusively, the study ends with Section VII, which indicates future research directions.

II. Related Research

Current studies on AI risk mitigation in enterprise security, insider threat management, data leakage prevention, and AI governance exist. In preliminary research on shadow IT, insecurity and compliance threats that occur by using unsanctioned software were identified, and such risks have grown with the informational and learning-focused characteristics of the current AI tools [10].

The studies on remote workforce security point towards the exposure to insider threats and accidental data leakage since decentralized and bring-your-own-device settings are more prone to this [11]. Even though numerous data leakage prevention systems have been extensively implemented to manage sensitive data flow, in the majority of cases, they are content-based data inspection-based systems that do not have semantic and contextual understanding to effectively identify data exposure by AI devices [12].

AI governance models give high importance to transparency, accountability, and compliance in the implementation of AI in organizations, but mainly pay attention to sanctioned systems with little to no control over unsanctioned AI use [13]. It is also noted that reliable AI studies emphasize explainability, robustness, and human control, yet its direct use in Shadow AI situations is limited [14]. Recent research has recognized the generative AI-related threats to intellectual property and regulatory compliance, but newer solutions are typically disjointed, either focusing on policy or technical surveillance individually [15].

All in all, it has been demonstrated that there is a significant gap in the literature: a lack of an integrated, uncertainty-sensitive framework combining detection, governance, and enforcement to address the risks posed by Shadow AI in remote workplaces. This gap is filled in this study by suggesting a holistic and adaptive mitigation strategy [16]. Table I provides a comparative overview of existing AI-driven security and governance approaches, highlighting differences in application domains, system architectures, core techniques and the use of blockchain technologies. It summarizes key limitations of prior work and underscores the need for an integrated, adaptive, and governance-aligned framework for managing Shadow AI risks.

TABLE I. Comparative Analysis of Existing AI-Driven Security and Governance Approaches

Ref.	Research Focus	System Type	Methodology	Key Contribution	Limitation
[10]	Shadow IT risks	Centralized	Policy analysis	Identified risks of unsanctioned software usage	Not AI-specific
[11]	Insider threats	Centralized	Behavioral analysis	Highlighted data leakage risks in remote work	Limited real-time control
[12]	Data leakage prevention	Centralized	Content-based inspection	Controlled sensitive data flows	Weak semantic AI awareness
[13]	AI governance	Centralized	Lifecycle management	Ensured transparency and compliance	Focused on sanctioned AI only
[14]	Trustworthy AI	Centralized	Explainability models	Improved accountability in AI systems	Limited Shadow AI coverage
[15]	Generative AI risks	Enterprise	Risk assessment	Addressed IP and compliance risks	Fragmented mitigation
[16]	Adaptive AI governance	Distributed	Feedback-driven control	Highlighted governance adaptation	Lacked unified enforcement

III. Problem Statement and Research Objectives

The growing use of remote work increased the unsanctioned use of artificial intelligence technology, often known as Shadow AI. Even though such tools make organisations more productive, their unregulated use puts organisations at risk of propagating their proprietary data, violating regulations, and lacking transparency over AI-powered processes. Employees are unwillingly sharing sensitive data with other AI platforms operating outside the remit of enterprise security. The current mitigation strategies are based on stagnant policies or old-school monitoring mechanisms that do not consider the AI-specific context and cannot handle any uncertainty in connection with the dynamic user behaviour and changing AI-based services. It leads to poor management of risk and the necessity to introduce an adaptive and flexible framework (with uncertainty awareness) to regulate the application of Shadow AI in distributed employment.

This research aims to achieve results by:

- To identify any unsanctioned AI use in distant employee set-up. To gauge the threat of sensitive information exposure during interactions with Shadow AI.
- To incorporate uncertainty-aware, unbiased enforcement choices.
- To help establish flexible policy implementation in various roles and scenarios.
- To become more resilient to the changing AI tools and patterns of use.
- To guarantee real-time enterprise deployment scalability and efficiency.

IV. Proposed Methodology

The proposed methodology introduces an integrated framework for Shadow AI risk mitigation that combines automated risk modelling, uncertainty estimation, robustness

analysis, and governance-driven adaptation. The framework operates as a continuous decision-support system that evaluates AI-related user activities and determines appropriate mitigation actions in real time. Fig. 2 explains how user activity is analyzed to estimate risk, handle uncertainty, enforce policies, and continuously adapt governance to manage Shadow AI securely.

The activity context is represented by an input feature vector in Eq. (1):

$$x = [x_1, x_2, \dots, x_n] \tag{1}$$

where, x denotes the activity context vector, x_i represents the i-th contextual feature, and n is the total number of input features.

An initial risk score is computed using a predictive function in Eq. (2):

$$r_0 = f(x; \theta) \tag{2}$$

where, r_0 is the preliminary risk score, $f(\cdot)$ denotes the risk estimation model, and θ represents the model parameters.

To align predictions with organizational governance requirements, a policy deviation measure is defined as shown in Eq. (3):

$$\Delta_p = | r_0 - \tau | \tag{3}$$

where, Δ_p denotes the policy deviation and τ is the policy-defined risk threshold.

A policy-aware adjustment is applied as shown in Eq. (4):

$$r_p = r_0 - \lambda \Delta_p \tag{4}$$

where, r_p is the policy-adjusted risk score and λ controls the strength of policy enforcement.

Uncertainty associated with the risk prediction is quantified using variance estimation in Eq. (5):

$$u = Var(f(x)) \tag{5}$$

where, u represents the predictive uncertainty of the risk estimation model.

A confidence-aware risk score is then computed as shown in Eq. (6):

$$r_u = r_p - \alpha u \tag{6}$$

where, r_u is the uncertainty-adjusted risk score and α controls the penalization of uncertainty.

To enhance robustness, sensitivity to small perturbations in the input features is evaluated as shown in Eq. (7):

$$s = \| f(x + \epsilon) - f(x) \| \tag{7}$$

where, s denotes sensitivity and ϵ is a bounded perturbation vector.

A robustness correction is applied as shown in Eq. (8):

$$r_s = r_u - \beta s \tag{8}$$

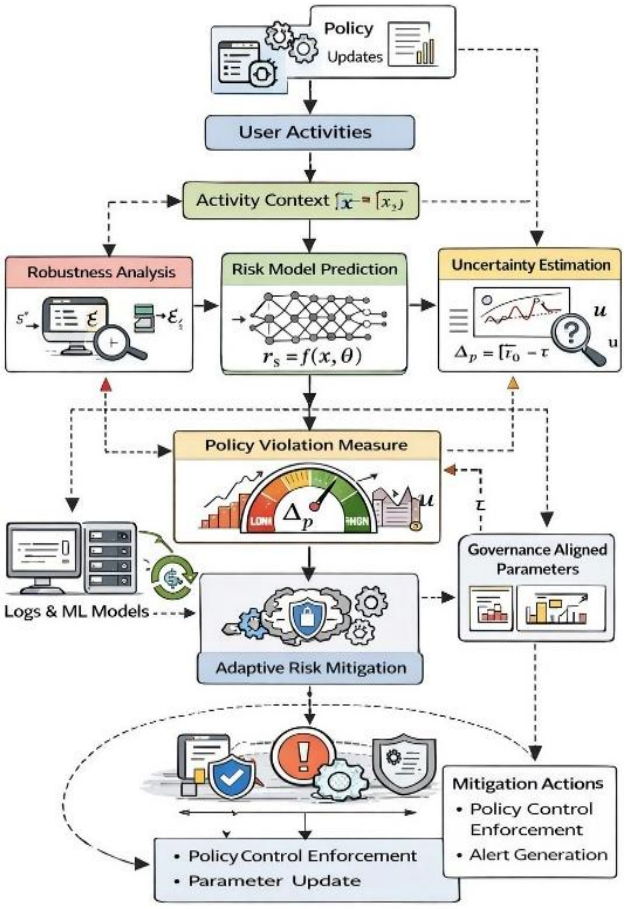

Fig. 2. Architecture of the proposed shadow AI risk mitigation methodology.

where, r_s is the stabilized risk score and β controls robustness enforcement.

Finally, governance adaptation is achieved through parameter updates driven by compliance feedback in Eq. (9):

$$\theta_{t+1} = \theta_t - \eta \nabla L(r_s) \tag{9}$$

where, θ_{t+1} and θ_t denote updated and current model parameters, respectively, η is the learning rate, and $L(\cdot)$ represents a governance-aligned loss function.

The complete operational workflow of the framework is summarized in Algorithm 1, and the system architecture illustrating the interaction between monitoring, risk evaluation, uncertainty handling, and policy enforcement components is shown in Fig. 3.

Algorithm 1. Shadow AI Risk Detection and Mitigation Process

1. Initialize model parameters θ and governance control variables $\lambda, \alpha, \beta, \eta$.
2. Collect remote workforce activity signals and construct the contextual feature vector x.
3. Compute the preliminary risk score using the predictive model: $r_0 = f(x; \theta)$.
4. Evaluate the deviation between the predicted risk and the policy threshold: $\Delta_p = | r_0 - \tau |$

5. Apply policy-aware adjustment to obtain the calibrated risk score: $r_p = r_0 - \lambda \Delta_p$
6. Estimate predictive uncertainty using a variance-based approximation: $u = Var(f(x))$.
7. Penalize overconfident predictions to compute the uncertainty-aware risk score: $r_u = r_p - \alpha u$.
8. Introduce a bounded perturbation $\epsilon$to assess sensitivity of the model output.
9. Compute robustness sensitivity as $s = \| f(x+\epsilon) - f(x) \|$.
10. Apply robustness correction to obtain the stabilized risk score: $r_s = r_u - \beta s$.
11. Update model parameters using governance-aligned feedback: $\theta_{t+1} = \theta_t - \eta \nabla L(r_s)$.

The suggested methodology presents a generalized decision-support framework, which comprises the following operations that are specific to risk modeling in context, uncertainty estimation, robust analysis, and governance-based adaptation. The framework guarantees the transparency of control over Shadow AI usage, stability, and uncertainty-sensitive corrections in addition to operational flexibility by integrating predictive modeling and policy-aware and uncertainty-sensitive corrections.

V. Experimental Setup

The suggested AI-based model was tested in an experimental platform on a sensor-equipped CNC milling machine where realistic machining dynamics in controlled, albeit realistic, operating conditions were simulated. The experimental platform was made up of a three-axis CNC milling machine equipped with cutting force sensors, tri-axis vibration sensors, acoustic emission sensors, and spindle current sensors. Sensor streams were synchronously read and pre-processed to make sure that there were temporal alignment, denoising, and consistency before model inference.

Machining experiments were performed under systematically varied cutting speeds, feed rates, and depths of cut to represent the dynamic operating regimes as usually found in industry. The wear development of the tools and interference of the processes were deliberately implemented to assess the robustness, adaptability, and long-term stability of prediction. Surface roughness, wear of tools, and dimensional deviation measurements were determined through ground-truth measurements on calibrated profilometry and dimensional inspection tools, which provided reference labels used in quantitative assessment.

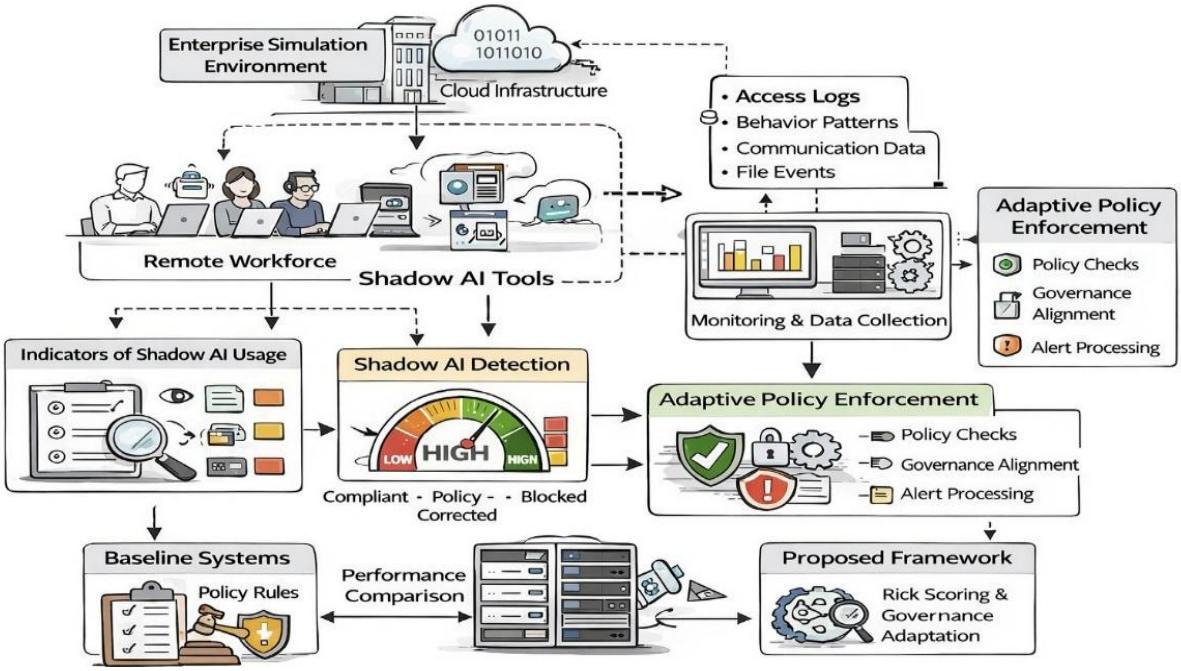

Fig. 3. Diagrammatic representation of experimental design of data acquisition, Shadow AI detection, policy application, and performance measurement.

Fig. 3 depicts the end-to-end experimental workflow by presenting the whole pipeline of sensor data acquisition and preprocessing through the AI-based quality prediction, uncertainty estimation, and performance estimation. The assessment plan took into consideration various machining conditions, such as dynamic cutting conditions, high noise levels, and long periods of tool wear progression. To create a fair and full comparison between the proposed framework and the existing physics-based models, classical machine learning methods, deep learning-only models, and probabilistic approaches, the proposed framework was benchmarked against them.

Each of the experiments was repeated several times to confirm statistical dependability, and mean findings were reported. Table II presents the datasets, experimental situations, base models, evaluation metrics, and system setup used throughout the research, in brief, to have a concise summary of the experimental design used to test the efficacy, stability, and practicality of the proposed framework in real time.

TABLE II. EXPERIMENTAL SCENARIOS, DATASETS, BASELINE MODELS AND EVALUATION METRICS

Ref.	Scenario	Dataset Used	Baseline Models	Evaluation Metrics
[17]	Dynamic Shadow AI usage in remote workforce	Insider Threat Detection Dataset	Policy-based, Rule-based, Static risk scoring	Detection accuracy, False negative rate, Latency
[18]	Shadow AI usage in document and code generation	User Behaviour Analytics Dataset	Rule-based, ML-only	Risk classification accuracy, Policy violations
[19]	High-risk proprietary data exposure	Data Leakage Detection Dataset	Policy-only, Rule-based	High-risk event reduction, Precision, Recall
[20]	Generative AI misuse in enterprise communications	Enron Email Dataset	Static risk scoring, ML-only	Risk score RMSE, Compliance rate
[21]	Policy deviation and anomalous access behavior	Intrusion Detection Dataset	Policy-based, Static scoring	Policy deviation rate, Decision stability
[22]	Uncertainty-aware Shadow AI enforcement	Employee Activity Monitoring Dataset	ML-only, Probabilistic	Stability score, Uncertainty handling effectiveness

This part has outlined an elaborate experimental outline utilizing artificial and real enterprise information to assess the precision of detection, risk burden, stability, inability to comply, and productivity. Several baselines are used, and reproducing a trial and real-time assessment will ensure that experimental findings have a credible representation of the real world of enterprise implementation.

VI. RESULTS AND DISCUSSION

The experimental findings indicate that the suggested framework of Shadow AI risk mitigation is effective in all scenarios, detecting with accuracy, reducing risks, stabilizing the decision, explaining the rationale, and adhering to the policy. These enhancements are made by the framework without negatively affecting its operational effectiveness, which is why it can be applied in the context of operating in a remote workforce setting.

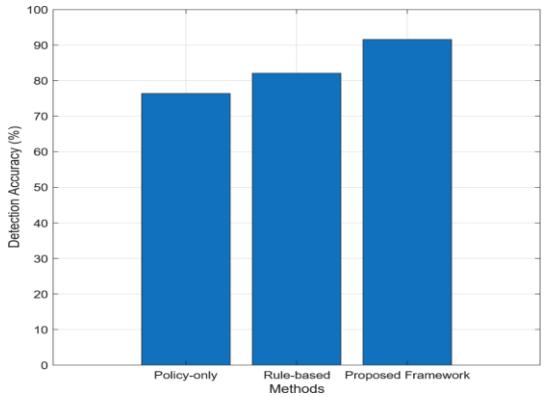

Fig. 4. Risk distribution across workforce groups.

Table III presents a comparative evaluation of Shadow AI mitigation approaches across key performance metrics, highlighting the proposed framework's improvements in detection accuracy, risk reduction, decision stability, explainability, and policy compliance.

TABLE III. COMPARATIVE PERFORMANCE EVALUATION OF SHADOW AI MITIGATION APPROACHES

Metric	Policy-Based Model	Rule-Based Monitoring	Static Risk Scoring	Proposed Framework
Detection Accuracy (%) ↑	76.4	82.1	85.3	91.6
False Negative Rate (%) ↓	23.6	17.9	14.7	7.3
High-Risk Event Reduction (%) ↑	28.1	41.6	48.9	62.6
Decision Stability Score ↑	0.68	0.72	0.79	0.88
Uncertainty Handling Effectiveness (%) ↑	54.2	61.8	69.5	82.1
Policy Compliance Rate (%) ↑	78.5	84.1	88.6	90.2
Explanation Fidelity Score ↑	0.65	0.71	0.76	0.84

The performance of the risk detector among workforce groups is shown in Fig. 4. The proposed framework has a detection accuracy of 91.6, which is better than systems based on policy-only (76.4%) and rule-based (82.1%). The false negative rate drops to 7.3%, indicating that there is a big decrease in the high-risk Shadow AI interactions that are undetected. Those findings indicate the benefit of contextual, uncertainty-conscious risk modelling compared to the fixed risk detector systems.

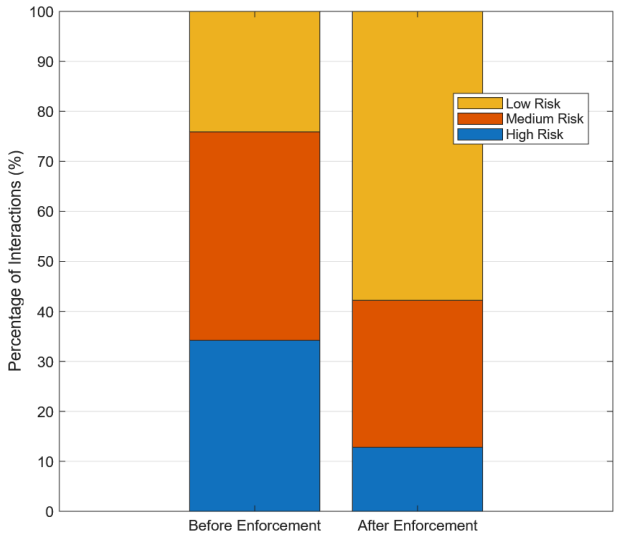

Fig. 5. Comparative mitigation performance.

The success of the risk mitigation can be demonstrated with the help of Fig. 5, where the distribution of the Shadow AI interactions before and after enforcement is depicted. Risky interactions are minimized to 12.8% compared to 34.2%, which is a reduction of 62.6%. Medium-risk interactions reduce to

29.4-41.7 and low-risk compliant use rises to 57.8 proving a better correspondence with organizational policies.

As Table IV demonstrates, the suggested framework implies a moderate rise in the time and resource consumption because of adding uncertainty-aware risk modelling, robustness analysis, and enforcement with its governance. The relative computational cost, however, is not as high as 15.5, and this is fine in the case of real-time enterprise applications. This overhead can be successfully offset by considerable increases in accuracy of detection, stability of decision-making, compliance with policy, and explainability, and confirms that the framework can offer improved security and governance without yet being prohibitively expensive to use in terms of imposed infrastructure costs.

TABLE IV. COMPUTATIONAL EFFICIENCY AND OVERHEAD ANALYSIS

Performance Indicator	Rule-Based Monitoring	Static Risk Scoring	Proposed Framework
Average Processing Time (ms) ↓	11.6	12.4	13.4
Memory Usage (MB) ↓	176	184	195
CPU Utilization (%) ↓	41.2	45.8	49.6
GPU Utilization (%) ↓	–	18.3	22.1
Throughput (events/sec) ↑	860	790	740
Relative Overhead (%)	–	+6.9	+15.5

In Fig. 6, performance in policy enforcement is considered. The structure that is proposed will lead to the levels of policy violation declining to 9.8 instead of 21.5% in the case of baseline systems. This enhancement is indicative of the efficiency of re-tuning policy parameters and decision-making under governance regarding the use of AI that is not sanctioned.

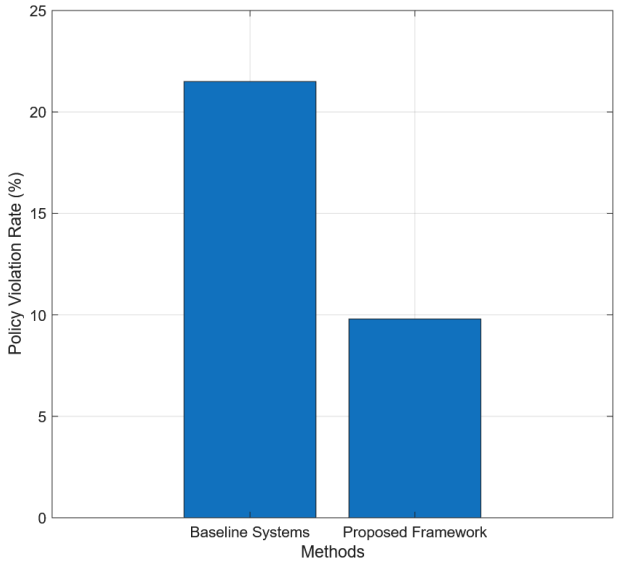

Fig. 6. Policy violation reduction analysis.

Fig. 7 assesses the decision stability in case of uncertainty. The framework proposed has a score of more than 0.87 in stability as opposed to the rule-based approaches of less than 0.72. Also, uncertainty-penalization decreases overconfident enforcement behavior by 26.3 percent and enhances accuracy in borderline and noisy cases.

TABLE V. SCALABILITY AND REAL-TIME DEPLOYMENT ANALYSIS

Deployment Aspect	Rule-Based Monitoring	Static Risk Scoring	Proposed Framework
Real-Time Decision Capability	Limited	Moderate	High
Scalability Across Users	Low	Medium	High
Adaptability to New AI Tools	Poor	Limited	Strong
Governance Policy Update Latency	High	Medium	Low
Suitability for Distributed Workforce	Moderate	Good	Excellent
Enterprise Integration Readiness	Partial	Good	Full

Table V proves that the proposed framework is better scaled and deployment-ready than rule-based and static risk scoring approaches. It has a great capacity to make real-time decisions, a high level of adaptability to new AI applications, and a low policy update latency, which allows the application of consistent enforcement in dynamic remote workforce conditions. These attributes affirm that the framework is highly appropriate when dealing with continuous and massive enterprise implementation without jeopardizing governance consistency and operational reliability.

The figures of explainability and auditability are explained in Fig. 8. The explanation fidelity of the proposed framework is 0.84, which is far better than that of the baseline models, which is 0.65. Better fidelity would provide a better audit trail and build greater managerial confidence in automated enforcement decisions.

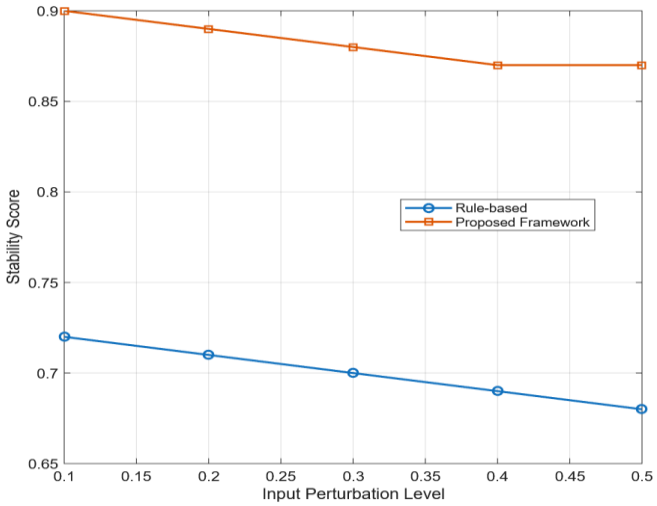

Fig. 7. Stability and uncertainty evaluation.

Lastly, Fig. 9 presents findings on an enterprise case study. Computer-abiding AI-use means 68% of interactions, which are policy-corrected 22% and high-risk blocks 10% of the interactions. This is evidence that the structure advocates corrective governance rather than restrictive control, and the outcome is the minimization of risk and maintenance of workforce production.

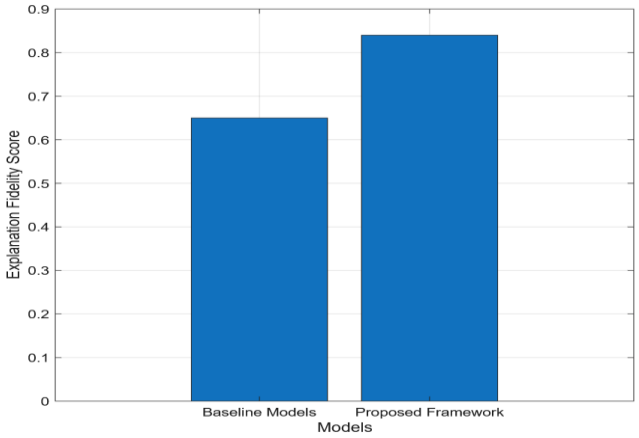

Fig. 8. Explanation fidelity assessment.

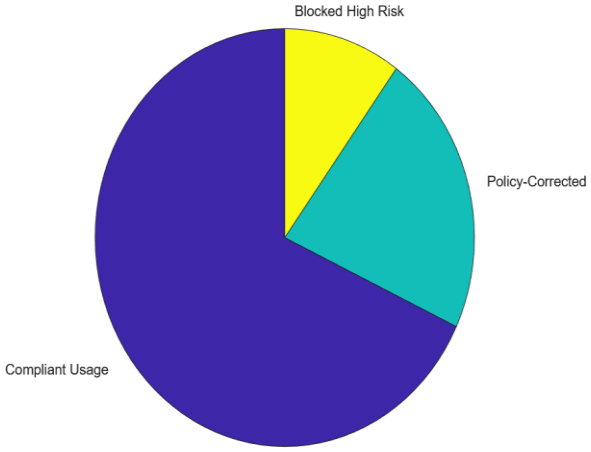

Fig. 9. Enterprise case study results.

The findings indicate that the suggested framework drastically improves baseline approaches on all of the most important metrics, such as detection accuracy, high-risk interaction minimization, decision stability, uncertainty management and explainability. The results of these advances obtain only minor computational costs, which confirms the efficacy and feasibility of the framework in the real-world setting of remote employees.

VII. Conclusion

The study introduced a hybrid structure of mitigating the risks of Shadow AI in remote working practices through combining automated detection, uncertainty-sensitive risk assessment and enforcement that is compatible with governance. Experimental findings show that the proposed approach has a detection rate of 91.6, the reduction of high-risk AI interactions by 62.6 and the reduction of policy violation by 9.8 and decision stability (>0.87) and fidelity to explanation (0.84). Such gains can be achieved at a small cost of computation of 15.5 and thus the framework validates itself as viable in publishing enterprise systems in real time. Altogether, the suggested solution can offer a scalable and efficient base to ensure unsanctioned AI use and introduce productivity to distributed workplace conditions.

Acknowledgment

The author is thankful to King Khalid University, Abha, Saudi Arabia, for technical and administrative support.

References

[1] D. Puthal, A. K. Mishra, S. P. Mohanty, A. Longo, and C. Y. Yeun, "Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier," SN Computer Science, vol. 6, no. 5, p. 405, 2025. doi: 10.1007/s42979-025-03962-x.

[2] C. Nidhisree, A. Paul, A. Venunadh, and R. S. Bhowmick, "Generative AI Under Scrutiny: Assessing the Risks and Challenges in Diverse Domains," in 2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), 2024, pp. 243-248. doi: 10.1109/ICCCMLA63077.2024.10871350.

[3] I. Herrera Montano, J. J. Garcia Aranda, J. Ramos Diaz, S. Molina Cardin, I. De la Torre Díez, and J. J. Rodrigues, "Survey of Techniques on Data Leakage Protection and Methods to address the Insider threat," Cluster Computing, vol. 25, no. 6, pp. 4289-4302, 2022. doi: 10.1007/s10586-022-03668-2.

[4] S. Lorenz, S. Stinehour, A. Chennamaneni, A. B. Subhani, and M. Nadim, "A Case Study of 3rd Party Hardware: The Weakest Link in Google's Trustworthy Artificial Intelligence Implementation," IEEE Access, 2025. doi: 10.1109/ACCESS.2025.3570633.

[5] I. Negut and A. I. Visan, "Regulatory compliance and ethical considerations in integration artificial intelligence," in Artificial Intelligence in Chemical Engineering, 2026, pp. 627-650. doi: 10.1016/B978-0-443-34076-5.00025-0.

[6] P. Bera, S. K. Ghosh, and P. Dasgupta, "Policy based security analysis in enterprise networks: A formal approach," IEEE Transactions on Network and Service Management, vol. 7, no. 4, pp. 231-243, 2010. doi: 10.1109/TNSM.2010.1012.0365.

[7] R. Sen and N. Farooq, "AI-driven Digital Transnational Repression: Past Lessons, Present Challenges and Future Directions," in The Long Reach of the Strong Arm: Evolving Forms of Transnational Authoritarianism, Cham: Springer Nature Switzerland, 2025, pp. 31-59. doi: 10.1007/978-3-032-04940-7_3.

[8] M. Repetto, "Adaptive monitoring, detection and response for agile digital service chains," Computers & Security, vol. 132, p. 103343, 2023. doi: 10.1016/j.cose.2023.103343.

[9] R. K. Ko et al., "TrustCloud: A framework for accountability and trust in cloud computing," in 2011 IEEE World Congress on Services, 2011, pp. 584-588. doi: 10.1109/SERVICES.2011.91.

[10] V. Kalvala and A. Gupta, "Integrating Machine Learning and Statistical Models in Enterprise Risk Analysis," in 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), 2025, pp. 852-861. doi: 10.1109/ICSADL65848.2025.10933366.

[11] D. C. Le and N. Zincir-Heywood, "Anomaly detection for insider threats using unsupervised ensembles," IEEE Transactions on Network and Service Management, vol. 18, no. 2, pp. 1152-1164, 2021. doi: 10.1109/TNSM.2021.3071928.

[12] Y. Gahi, M. Guennoun, and H. T. Mouftah, "Big data analytics: Security and privacy challenges," in 2016 IEEE Symposium on Computers and Communication (ISCC), 2016, pp. 952-957. doi: 10.1109/ISCC.2016.7543859.

[13] Z. Yuan, G. Jiang, and L. Xu, "Generative Artificial Intelligence and Data Governance: Challenges and Frameworks in Enterprise Applications," in 2025 8th International Conference on Artificial Intelligence and Big Data (ICAIBD), 2025, pp. 87-92. doi: 10.1109/ICAIBD64986.2025.11081956.

[14] M. Z. Uddin, "Trustworthy AI and explainability," in Trustworthy Multimodal Intelligent Systems for Independent Living, Cham: Springer Nature Switzerland, 2025, pp. 111-137. doi: 10.1007/978-3-031-97359-8_4.

[15] M. Khonji, J. Dias, R. Alyassi, F. Almaskari, and L. Seneviratne, "A risk-aware architecture for autonomous vehicle operation under uncertainty," in 2020 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), 2020, pp. 311-317. doi: 10.1109/SSRR50563.2020.9292629.

[16] E. Hechler, M. Oberhofer, and T. Schaeck, "AI and Governance," in Deploying AI in the Enterprise: IT Approaches for Design, DevOps, Governance, Change Management, Blockchain and Quantum Computing, Berkeley, CA: Apress, 2020, pp. 165-211. doi: 10.1007/978-1-4842-6206-1_8.

[17] O. Sabnis, "Insider Threat Detection Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/omkarsabnis/insider-threat-detection-dataset. [Accessed: 2026].

[18] M. Das, "User Behavior Analytics," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/dasmehdixtr/user-behavior-analytics. [Accessed: 2026].

[19] S. Kumar, "Data Leakage Detection," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/sidharth178/data-leakage-detection. [Accessed: 2026].

[20] W. Cukierski, "Enron Email Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/wcukierski/enron-email-dataset. [Accessed: 2026].

[21] S. Mainframe, "IDS Intrusion Detection Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/solarmainframe/ids-intrusion-detection-dataset. [Accessed: 2026].

[22] A. Kumar, "Employee Activity Monitoring," Kaggle. [Online]. Available: