PaperGuard AI

Prof. Dhanshri Choudhari[1] Vaishnavi Bhosale[2] Dhanashri Ekke[3] Akshara Rangapure[4] Tejashree Shinde[5]

[1,2,3,4,5]Department of Artificial Intelligence and Machine Learning

[1,2,3,4,5]Zeal Polytechnic, Pune, Maharashtra, India

Abstract — PaperGuard is a proposed AI-assisted security framework intended to reduce the risk of examination paper leakage before sensitive content is released. It combines role-based access control, authenticated access, multi-factor authentication, biometric verification, an AI risk engine, AES-256 double encryption, an encrypted cloud vault, forensic watermarking, and secure kiosk-mode display. The AI layer is designed to learn normal access patterns, analyze access telemetry, identify unusual behavior, explain risk factors, and prioritize alerts. A risk score above the proposed 30% threshold triggers review rather than automatic release. Super-Admin approval remains a required human control. The paper describes the proposed architecture and end-to-end workflow from upload to controlled examination access. The design is a project proposal; its effectiveness requires implementation and evaluation.

Keywords: PaperGuard, Examination Security, Anomaly Detection, Risk Scoring, AES-256, Forensic Watermark, Access Control

I. Introduction

Examination papers contain sensitive information, and unauthorized access before an examination can undermine fairness and trust. A security design must protect the document itself while also controlling who can request access, when access is allowed, and how decisions are recorded. PaperGuard proposes a centralized framework in which AI assists security teams by identifying unusual access behavior and prioritizing requests for review.

The main idea is to combine technical safeguards with human oversight. Encryption protects stored content; identity and role checks restrict access; AI analyzes access telemetry; and a Super-Admin makes the final approval decision. The framework also proposes forensic watermarking and secure kiosk-mode display to support accountability after release.

II. Proposed System

A. Security Triad

PaperGuard organizes its controls into three cooperating components:

1) Central Command: The Super-Admin reviews AIgenerated recommendations and approves or denies sensitive access. Reviewed events can be used as feedback to improve later recommendations.
2) AI Risk Engine: The proposed engine learns baseline behavior for roles, teams, and examination windows. It continuously analyzes telemetry, detects adaptive anomalies, identifies explainable risk factors, and assigns a risk score from 0–100%.
3) Dynamic Protection: AES-256 double encryption, an encrypted cloud vault, and invisible forensic watermarks protect papers. AI prioritizes alerts and recommends the next protection step, while the human approval requirement is retained.

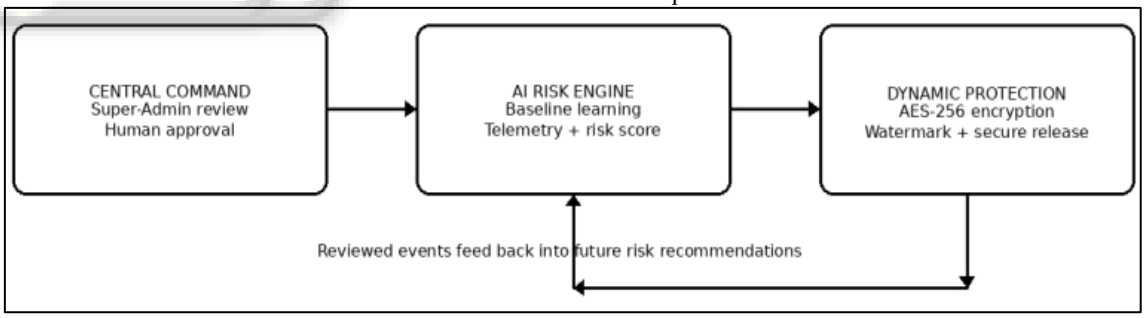


Fig. 1: PaperGuard security triad and feedback loop.

B. Zero-Trust Access

The proposed zero-trust principle is "Never trust. Always verify." Each request is authenticated, checked against role-based access control (RBAC), scored by the AI layer, considered by the authorized reviewer, and logged. Permissions should be limited to the minimum required for each role. AI recommendations support the review process but do not replace the Super-Admin.

III. End-to-End Workflow

The proposed workflow begins when the Super-Admin uploads an examination paper. The document is protected using AES-256 double encryption and stored in an encrypted cloud vault. Before a user can request the paper, the system requires official login with multi-factor authentication (MFA) and biometric verification. The AI risk engine compares access behavior with its learned baseline and evaluates relevant telemetry.

If the risk score exceeds the proposed 30% threshold, the system prioritizes an alert and recommends review.

The Super-Admin then decides whether to approve access. A denied request is blocked, and the reviewed event can feed back into the AI process. For an approved request, the paper is scheduled for decryption 15 minutes before the examination, a forensic watermark is embedded, and the content is displayed in secure kiosk mode. The examination

All rights reserved by www.ijsrd.com

is then conducted, and review feedback supports future recommendations.

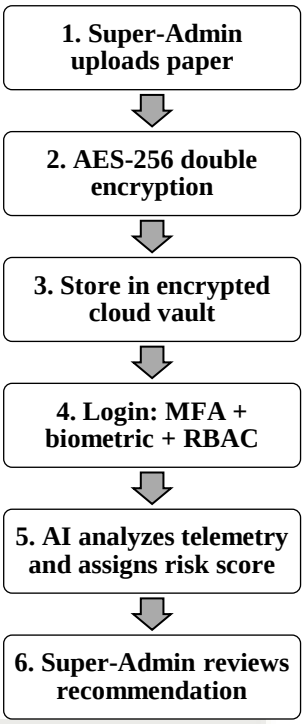


Fig. 2: Proposed end-to-end secure paper handling workflow.

IV. Risk Analysis and Human Oversight

The risk score is intended to help staff focus on unusual or potentially high-impact requests. Explainable risk factors are important because reviewers need a reason for an alert rather than an unexplained score. A score above 30% is a proposed escalation threshold from the project design, not a validated result. The final decision remains with the Super-Admin, reducing dependence on an automated decision alone.

The framework also proposes continuous learning from reviewed access events. Such feedback should be used with appropriate governance and retained evidence so that adaptation does not hide the reason for previous decisions. The project presentation describes the intended design but does not provide measured accuracy, false-positive rates, penetration-test results, or deployment outcomes.

V. Feasibility and Expected Impact

The proposal describes a lightweight AI service for baseline learning, telemetry analysis, adaptive anomaly detection, and alert prioritization. Its components can be integrated around an encrypted document vault and a controlled release workflow. Practical deployment would require secure identity integration, access logging, key management, model monitoring, backup and recovery procedures, and testing under realistic examination conditions.

Expected benefits include earlier identification of suspicious access, traceable approval decisions, protection of stored and released papers, and evidence to support investigations. The proposed watermark is intended to help link a leaked photograph to its originating device and user; this capability would need to be tested and validated in implementation. Overall, PaperGuard aims to turn separate controls into a centralized and explainable examination security process.

A. Audit Trail and Release Control

For each request, the proposed audit trail should make it possible to reconstruct the decision sequence: identity verification, role and permission checks, telemetry assessment, risk score, reviewer decision, scheduled decryption, watermarking, and display in kiosk mode. Recording the sequence provides a basis for post-event review and helps distinguish a denied request from an approved release. The project concept calls for a traceable paper release; the exact log format and retention period remain to be defined during implementation.

B. Operational Boundaries

AI-assisted anomaly detection can identify behavior that differs from a learned baseline, but unusual behavior is not automatically proof of malicious intent. The system should therefore present understandable risk factors and allow an authorized reviewer to consider context. Baselines may change as staff roles, examination schedules, and normal working patterns change. Model updates should be reviewed and monitored so that legitimate changes do not create excessive alerts or weaken the intended controls.

The proposed workflow also depends on reliable coordination between authentication, risk scoring, approval, vault access, timed decryption, and kiosk display. Testing should verify that these components use the same examination schedule and authorization state. The system should avoid exposing plaintext during upload, storage, or unauthorized requests, and should preserve an auditable record when a service fails or an access request is rejected.

VI. Implementation Considerations

A practical implementation should separate document storage, identity verification, risk analysis, approval, and release into auditable modules. The upload module should verify that the uploader is authorized before accepting a paper. The storage layer should retain only encrypted content and restrict access to decryption keys. The authentication layer should validate official credentials and the configured MFA and biometric checks. The AI service should receive only the access events needed for risk analysis, while the approval module records the reviewer, decision, timestamp, and reason.

The system should fail safely. If identity verification, the risk service, or the approval service is unavailable, the paper should not be released automatically. Access logs should be protected from unauthorized modification and should capture both allowed and denied requests.

Administrative permissions should be reviewed periodically, and test environments should use nonsensitive sample papers. These are implementation considerations derived from the proposed security goals; the presentation does not specify a finalized deployment platform or key-management design.

All rights reserved by www.ijsrd.com

A. Proposed Component Responsibilitie

The Super-Admin interface is the control point for uploading papers and reviewing sensitive access decisions. The authentication component verifies the identity of a requester, while RBAC checks whether that identity is permitted to request the particular examination paper. The risk engine provides a recommendation based on the request context and learned access baseline. The encrypted vault stores the protected document, and the release component coordinates decryption timing, watermark embedding, and kiosk-mode display. Separating these responsibilities makes the workflow easier to test and audit.The boundaries between components should be explicit. An AI score should not directly expose a paper, and a successful login should not by itself grant access to every examination document. A release should require the relevant permission and recorded approval. Similarly, the forensic watermark is an accountability measure and does not replace encryption, authentication, or approval. These distinctions follow the proposed security triad, in which central command, AI analysis, and dynamic protection provide complementary rather than interchangeable defenses.

B. Limitations of the Current Proposal

The supplied presentation describes the architecture and intended benefits but does not provide implementation code, training data, a selected anomaly-detection algorithm, a defined risk-score formula, or measured results. It also does not specify the encryption-key lifecycle, cloud provider, biometric service, watermarking technique, or integration details for the secure kiosk. These decisions must be documented before the system can be considered production-ready. In particular, the proposed 30% risk threshold should be treated as an initial design setting and calibrated through testing, not as an established security boundary.

VII. Evaluation Plan

This paper presented a survey-informed architecture for authorization-aware multi-agent RAG. Its central requirement is to restrict the retrieval candidate set using current authorization decisions before content reaches rerankers, agents, or a language model. The design combines access-control policy, least-privilege delegation, provenance, and audit logging. Future work is to implement the prototype and measure security, relevance, and latency against a retrieve-then-filter baseline.

VIII. Conclusion and Future Scope

PaperGuard proposes a layered approach to examination-paper security, combining encryption, restricted access, AI-assisted risk scoring, human approval, timed decryption, forensic watermarking, and secure kiosk-mode display. Its central design principle is that AI should help identify and prioritize risk while an authorized human retains the release decision. Future work should implement a prototype, define measurable security requirements, test anomaly detection on representative access logs, assess false alerts, validate watermark traceability, and conduct security and usability evaluations before operational use.

Acknowledgment

The authors acknowledge the project presentation titled “PaperGuard AI,” which supplied the proposed system concepts and workflow.

References

[1] Sandhu, R. S., Coyne, E. J., Feinstein, H. L., and Youman, C. E. (1996). “Role-Based Access Control Models,” IEEE Computer, 29(2), pp. 38–47.
[2] Chandola, V., Banerjee, A., and Kumar, V. (2009). “Anomaly Detection: A Survey,” ACM Computing Surveys, 41(3).
[3] National Institute of Standards and Technology (2001). “Advanced Encryption Standard (AES),” FIPS PUB 197.
[4] Rose, S., Borchert, O., Mitchell, S., and Connelly, S. (2020). “Zero Trust Architecture,” NIST Special Publication 800-207.

All rights reserved by www.ijsrd.com