

Deepfake-Aware Recruitment: An Enhanced Framework for Preventing Synthetic-Identity Insider Threats

Shaikh Huzaifa Saalim¹, Prof. M. N. Pimple², Prof. A. R. Kardile³, Dr. H. B. Jadhav⁴

Student, Computer Department¹

Assistant Professor, Computer Department^{2,3}

Head of Department, Computer Department⁴

Adsul's Technical Campus, Ahilyanagar, Maharashtra, India

Abstract: Remote recruitment has expanded access to talent but has also created opportunities for attackers to present fabricated or manipulated identities during interviews and onboarding. Deepfake audio and video, forged documents, and compromised online personas can undermine conventional identity checks. This paper examines deepfake-enabled synthetic-identity infiltration as a personnel-security and cybersecurity problem. Based on analysis of the supplied research paper and the risk-management practices it discusses, it identifies gaps in current hiring workflows and proposes a layered Deepfake-Aware Hiring Assurance Framework. The framework combines role-based risk classification, independent identity and credential verification, consent-based interview media analysis, live challenge-response checks, human review, least-privilege access, and continuous monitoring. Existing-process, proposed-system, and operational workflow diagrams are presented with explanations. The approach is conceptual: no prototype, dataset, or measured detection accuracy is claimed. Media-forensics tools are treated as supporting signals rather than definitive proof, with transparent governance, privacy safeguards, and human oversight needed for fair implementation.

Keywords: deepfake detection, synthetic identity, insider threat, remote recruitment, cybersecurity, risk management, identity verification

I. INTRODUCTION

Deepfake systems can generate or alter audio, video, and images so that synthetic content appears authentic. These capabilities support legitimate activities such as dubbing, accessibility, education, and creative production; however, they can also be misused to impersonate real people or create convincing false personas. The security concern is especially significant when a digital identity is used to obtain trusted access to an organization rather than merely to mislead a public audience. Remote hiring relies heavily on digital applications, video interviews, electronic documents, and online references. Each channel can be manipulated, and apparently consistent information across channels may still originate from a single fabricated persona. Reports discussed in the source paper describe deepfake-assisted impersonation in financial fraud, public-sector communications, and attempts to obtain employment at technology organizations. Such cases illustrate why identity assurance must be part of the enterprise security lifecycle. This paper reframes the source paper's recommendations into a practical, layered process that connects recruitment controls with cybersecurity operations. It does not assume that every remote applicant is suspicious or that a detector can reliably classify all synthetic media. Instead, it proposes proportionate checks based on role sensitivity, independent corroboration, documented review, and controlled access.



II. PROBLEM STATEMENT

Conventional recruitment checks may verify documents and references without adequately testing whether the person appearing in a remote interview is the legitimate applicant. Deepfake detectors can produce false positives and may perform differently across recording conditions and demographic groups. The problem is therefore not simply to detect manipulated pixels or audio; it is to establish trustworthy identity and authorization decisions when individual evidence sources may be unreliable.

III. OBJECTIVES AND CONTRIBUTIONS

The objectives are to:

- Summarize the existing hiring and risk-management approach described in the source paper;
- Identify principal control gaps for synthetic-identity threats;
- Design an enhanced framework and operational workflow;
- Discuss implementation, privacy, fairness, and evaluation requirements.

The contribution is a conceptual integration of identity proofing, multimodal media analysis, human review, and post-hire monitoring into a single recruitment assurance process.

IV. BACKGROUND AND RELATED WORK

Deepfake generation has advanced alongside deep-learning methods and the availability of large media datasets. Yet detection remains a moving target: compression, lighting, facial occlusion, speech variation, and newer synthesis techniques can reduce the reliability of automated classifiers. Public benchmarking efforts, including the Deepfake Detection Challenge and NIST's Open Media Forensics Challenge, demonstrate the importance of testing models on held-out and changing data rather than relying on a single headline accuracy value. The source paper reviews insider threats, in which an individual with legitimate access may misuse that access or enable another actor to do so. A malicious hire can potentially reach source code, customer information, credentials, or internal infrastructure. The threat is therefore a combination of identity deception, personnel risk, and technical access—not solely a media-authenticity problem.

The NIST Risk Management Framework provides a lifecycle for preparing, categorizing, selecting, implementing, assessing, authorizing, and monitoring security controls. NIST SP 800-53 includes controls relevant to personnel security, least privilege, audit, training, incident handling, and continuous monitoring. These controls provide a foundation, but organizations must translate them into concrete recruitment steps and assign clear responsibility across Human Resources (HR), identity operations, and security teams.

V. EXISTING SYSTEM AND ITS LIMITATIONS

The existing process represented in Fig. 1 is a generalized interpretation of the remote-hiring process described in the source paper, not a claim that all organizations follow an identical procedure. A candidate submits an application, attends a remote interview, undergoes document and reference checks, receives a hiring decision, and is provisioned with an account. Where verification is primarily document-based or visual, a synthetic identity may pass through multiple stages without independent corroboration.

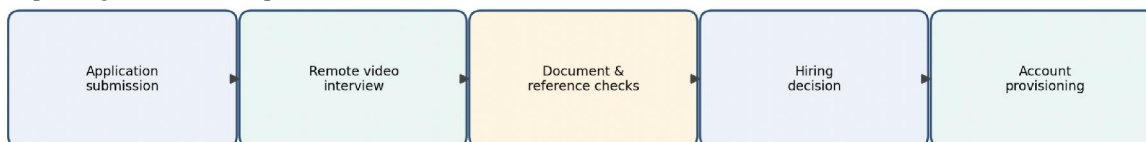


Fig. 1. Generalized existing remote recruitment and access process.



A. Limitations in the Existing Approach

- Single-channel reliance: a convincing video call or apparently valid document may be treated as stronger evidence than it warrants.
- Fragmented responsibility: HR may own candidate experience while security teams own technical risk, leaving no clearly accountable owner for high-risk identity checks.
- Late-stage control placement: access restrictions and monitoring may begin only after a candidate has been hired and provisioned.
- Detector overreliance: a media classifier can be wrong, and a low-risk result cannot prove that a person's identity is genuine.
- Privacy and fairness concerns: unbounded social-media screening or opaque automated rejection can introduce bias, collect irrelevant information, and reduce candidate trust.

VI. PROPOSED DEEPPAKE-AWARE HIRING ASSURANCE FRAMEWORK

The proposed framework in Fig. 2 uses multiple independent evidence sources and places a human decision-maker at the center of consequential hiring decisions. Controls are scaled to the sensitivity of the role. A position with access to production infrastructure or confidential intellectual property may require stronger corroboration than a role with no privileged access. The framework should be implemented consistently and in accordance with applicable employment, privacy, and data-protection requirements.

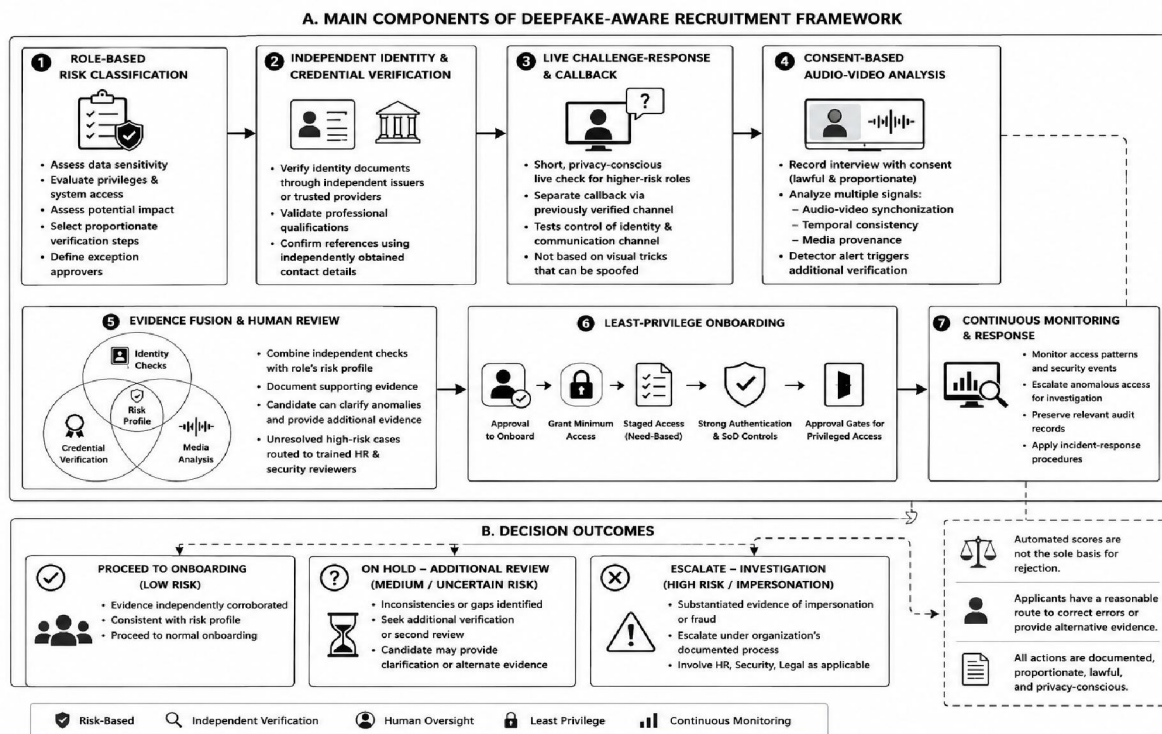


Fig. 2. Proposed layered framework for deepfake-aware recruitment assurance.

A. Main Components

Role-based risk classification: Classify the position according to data sensitivity, privileges, system access, and potential impact. Use this classification to select proportionate verification steps and define who approves exceptions.



Independent identity and credential verification: Validate identity documents and professional qualifications through appropriate independent issuers or trusted providers. Confirm references through contact details obtained independently rather than relying only on information supplied in the application.

Live challenge-response and callback: For higher-risk roles, use a short, accessible, privacy-conscious live check or a separate callback through a previously verified channel. The goal is to test control of the claimed identity and communication channel, not to rely on a visual trick that can itself be spoofed.

Consent-based audio-video analysis: Where lawful and proportionate, analyze interview recordings with multiple signals, such as audio-video synchronization, temporal consistency, and media provenance. A detector alert should trigger additional verification, not automatic rejection.

Evidence fusion and human review: Combine independent checks with the role's risk profile. Document supporting evidence, allow the candidate to clarify anomalies, and route unresolved high-risk cases to trained HR and security reviewers.

Least-privilege onboarding: If approved, grant only the access needed initially. Use staged access, strong authentication, separation of duties, and approval gates before granting privileged permissions.

Continuous monitoring and response: After onboarding, monitor access patterns and security events under approved policy. Escalate anomalous access for investigation, preserve relevant audit records, and apply established incident-response procedures.

B. Decision Outcomes

The framework supports three outcomes. A low-risk case with independently corroborated evidence can proceed to normal onboarding. A case with unresolved inconsistencies should be placed on hold for additional verification or a second review. A case with substantiated evidence of impersonation should be escalated under the organization's documented process. Automated scores must not be the sole basis for rejection, and applicants should have a reasonable route to correct errors or provide alternative evidence.



VII. OPERATIONAL WORKFLOW

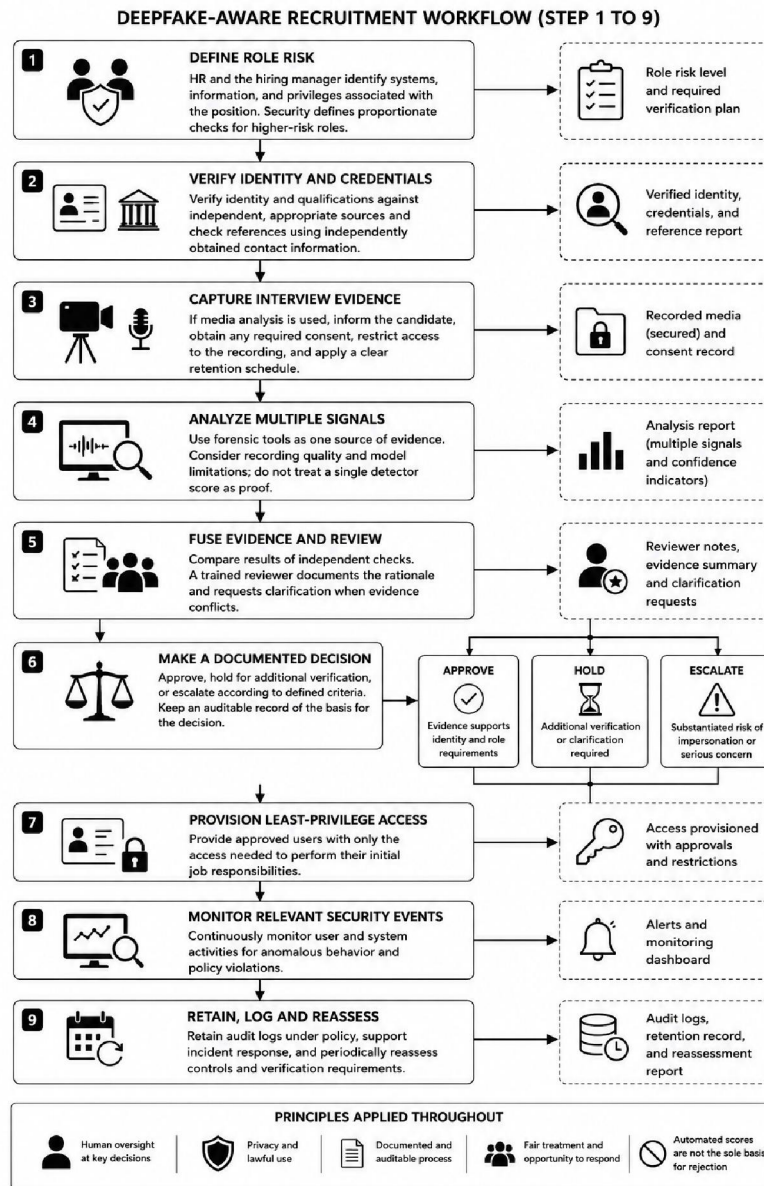


Fig. 3. Step-by-step operational workflow for deepfake-aware recruitment.

It starts with role classification and ends with monitored, auditable access. The workflow is designed to integrate with existing recruitment processes rather than replace HR judgment or standard security controls.

A. Workflow Explanation

Step 1 — Define role risk: HR and the hiring manager identify systems, information, and privileges associated with the position. Security defines proportionate checks for higher-risk roles.



Step 2 — Verify identity and credentials: Verify identity and qualifications against independent, appropriate sources and check references using independently obtained contact information.

Step 3 — Capture interview evidence: If media analysis is used, inform the candidate, obtain any required consent, restrict access to the recording, and apply a clear retention schedule.

Step 4 — Analyze multiple signals: Use forensic tools as one source of evidence. Consider recording quality and model limitations; do not treat a single detector score as proof.

Step 5 — Fuse evidence and review: Compare results of independent checks. A trained reviewer documents the rationale and requests clarification when evidence conflicts.

Step 6 — Make a documented decision: Approve, hold for additional verification, or escalate according to defined criteria. Keep an auditable record of the basis for the decision.

Steps 7–9 — Provision, monitor, and reassess: Provide approved users with least-privilege access, monitor relevant security events, retain audit logs under policy, and periodically reassess controls.

VIII. IMPLEMENTATION AND GOVERNANCE

A. Roles and Accountability

HR should own candidate communication and the fairness of recruitment decisions. Identity or compliance personnel should validate identity and credentials. Security architects define the control design; security engineers configure verification, logging, and access controls; and an authorized manager accepts residual risk. A named recruitment-security liaison should coordinate these responsibilities and ensure that high-risk cases do not fall between HR and IT.

B. Privacy, Fairness, and Security Safeguards

Organizations should collect only information necessary for the stated purpose, disclose how interview media will be used, limit access to recordings, and delete them according to a documented retention schedule. Social-media review, if used, should be legally reviewed, job-related, and consistent across candidates. Media-analysis systems should be assessed for error rates, demographic performance differences, accessibility impacts, and susceptibility to adversarial manipulation. Candidates need a human review path when an automated tool raises a concern.

C. Phased Deployment

A practical deployment can proceed in three phases. During months 0–3, conduct a role-based risk assessment, map the current recruitment process, define decision ownership, and approve privacy and retention rules. During months 3–9, pilot independent verification, live challenge-response, and media analysis for selected high-risk roles; measure false positives, false negatives where ground truth is available, processing time, and candidate complaints. During months 9–18, expand validated controls, train HR and security teams, audit outcomes, and periodically review tools as deepfake methods evolve. These intervals are a proposed planning schedule, not measured implementation results.

IX. EVALUATION PLAN

Because the framework is conceptual and no prototype or experimental dataset was supplied, this paper does not report detection accuracy or claim empirical validation. A future evaluation should use a legally obtained, consented dataset containing genuine and manipulated interview samples, with varied lighting, compression, accents, devices, and accessibility conditions. Identity verification should be evaluated separately from media classification so that detector performance is not mistaken for overall identity assurance.



Useful measures include:

- false-positive rate, to estimate how often genuine applicants are flagged;
- false-negative rate, to estimate how often manipulated samples pass the detector;
- verification completion and escalation rates;
- time and cost per candidate;
- reviewer agreement;
- performance differences across relevant candidate groups; and
- post-onboarding security incidents.

Evaluation should include independent test data and periodic re-testing to detect model drift. Thresholds should reflect role risk and the consequences of errors.

X. DISCUSSION

The main strength of the proposed framework is defense in depth: an attacker would need to defeat several checks designed to be independent, while a single detector failure need not determine the outcome. The framework also connects pre-employment checks to least-privilege provisioning and ongoing monitoring, recognizing that hiring is one part of the broader access lifecycle.

There are trade-offs. Additional verification can slow recruitment, create cost, and inconvenience legitimate candidates. Automated media tools can be inaccurate, and no single test guarantees authenticity. A strict requirement for in-person attendance may exclude qualified candidates or be impractical across geographies. Organizations should choose proportionate controls, provide accessible alternatives, and use risk-based escalation rather than treating remote work itself as evidence of wrongdoing.

XI. CONCLUSION

Deepfake-enabled synthetic identities create a challenge at the intersection of recruitment, identity assurance, and cybersecurity. Existing risk-management frameworks provide important controls for personnel screening, least privilege, audit, and continuous monitoring, but these controls need to be translated into explicit hiring procedures. This paper proposed a layered assurance framework combining role-based risk classification, independent identity and credential checks, consent-based media analysis, live verification, human review, restricted initial access, and continuous monitoring. The diagrams and workflow provide an implementation-oriented view of how these controls can fit together. The framework remains a conceptual proposal and should be piloted and evaluated before being treated as proven. Effective deployment depends on clear accountability, candidate privacy, fairness, and regular reassessment.

REFERENCES

- [1] Á. F. Gambin, A. Yazidi, A. Vasilakos, H. Haugerud, and Y. Djenouri, "Deepfakes: current and future trends," *Artificial Intelligence Review*, vol. 57, no. 3, 2024, doi: 10.1007/s10462-023-10679-x.
- [2] J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, "Deepfakes: Trick or treat?," *Business Horizons*, vol. 63, no. 2, pp. 135–146, 2020, doi: 10.1016/j.bushor.2019.11.006.
- [3] H. Chen and K. Magramo, "Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'," *CNN*, Feb. 4, 2024. [Online]. Available: <https://www.cnn.com/2024/02/04/asia/deepfake-cfo-hong-kong-intl-hnk/index.html>
- [4] J. Tidy, "Firm hacked after accidentally hiring North Korean cyber criminal," *BBC News*. [Online]. Available: <https://www.bbc.com/news/articles/ce8vedz4yk7o>
- [5] E. Gordenker, "False Face: Unit 42 Demonstrates the Alarming Ease of Synthetic Identity Creation," *Palo Alto Networks Unit 42*. [Online]. Available: <https://unit42.paloaltonetworks.com/north-korean-synthetic-identity-creation/>
- [6] "Deepfake Detection Challenge Dataset," *Meta AI*. [Online]. Available: <https://ai.meta.com/datasets/dfdc>



- [7] "Open Media Forensics Challenge," National Institute of Standards and Technology. [Online]. Available: <https://www.nist.gov/itl/iad/mig/open-media-forensics-challenge>
- [8] Joint Task Force, "Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy," NIST SP 800-37 Rev. 2, 2018, doi: 10.6028/NIST.SP.800-37r2.
- [9] Joint Task Force, "Security and Privacy Controls for Information Systems and Organizations," NIST SP 800-53 Rev. 5, 2020, doi: 10.6028/NIST.SP.800-53r5.
- [10] Cybersecurity and Infrastructure Security Agency, "Defining Insider Threats." [Online]. Available: <https://www.cisa.gov/topics/physical-security/insider-threat-mitigation/defining-insider-threats>
- [11] H. Guan, Y. Lee, L. Diduch, and I. Ghorbanian, "OpenMFC 2020–2021 Results," NIST, Dec. 8, 2021.
- [12] G. M. Dimitri, "A Short Survey on Deep Learning for Multimodal Integration: Applications, Future Perspectives and Challenges," Computers, vol. 11, no. 11, p. 163, 2022, doi: 10.3390/computers11110163.

