

# International Journal Research Publication Analysis

Page: 01-08

---

## HUMAN FACTORS, SYSTEM VULNERABILITIES, AND AI-DRIVEN THREATS IN ORGANIZATIONAL DATA BREACHES: A CONTEMPORARY REVIEW

---

**\*Dhanashri Muratkar, Dhanashri C. Kherkar, Prof. Narendra J. Padole**

MCA, Degree College of Physical Education (H.V.P. Mandal), Amravati, Maharashtra, India.

---

Article Received: 30 August 2026

Article Revised: 19 September 2026

Published on: 09 October 2026

**\*Corresponding Author: Dhanashri Muratkar**

MCA, Degree College of Physical Education (H.V.P. Mandal), Amravati, Maharashtra, India.

DOI: <https://doi-doi.org/101555/ijrpa.8944>

---

### ABSTRACT

Data breaches are no longer well explained by a simple choice between “human error” and “technical vulnerability.” Recent evidence shows that the dominant risk is increasingly socio-technical: users, administrators, identities, software flaws, third-party services, cloud configurations, and AI-enabled attack methods interact to create breach paths. This review re-examines the causes of organizational data breaches using recent industry evidence and a structured comparison of human, technical, and emerging AI-related factors. The analysis updates the earlier human-versus-system framing by incorporating vulnerability exploitation, credential abuse, social engineering, supply-chain exposure, shadow AI, and AI-assisted attacks. Verizon’s 2026 Data Breach Investigations Report indicates that vulnerability exploitation became the leading initial access vector, while human involvement and third-party exposure remain substantial. IBM’s 2026 Cost of a Data Breach research further reports a sharp rise in AI-enabled attacks and record global breach costs. The paper argues that organizations should move from isolated awareness training or patching programs toward integrated controls that combine identity security, rapid vulnerability remediation, secure configuration, third-party governance, AI usage controls, and continuous detection. The resulting framework is suitable for organizations seeking a practical, risk-based breach-prevention strategy under rapidly changing attack conditions.

**KEYWORDS:** Data Breach, Cybersecurity, Human Error, System Vulnerability, Artificial Intelligence, Social Engineering, Credential Abuse, Vulnerability Exploitation, Third-Party Risk, Zero Trust.

## I. INTRODUCTION

Digital transformation has increased the amount of sensitive information handled through cloud platforms, web applications, mobile services, APIs, third-party integrations, and AI-enabled systems. This expansion has improved organizational efficiency, but it has also increased the number of potential paths through which data can be exposed. Earlier breach analysis often separated causes into two broad categories: human error and system vulnerability. That distinction remains useful for basic classification, yet current incidents increasingly show that the categories overlap.

The original version of this study focused on the relative contribution of human error and technical weakness. Its central insight remains valid: many breaches involve both. A phishing message becomes more damaging when multifactor authentication is absent; an unpatched edge device remains vulnerable because patching was delayed; an exposed cloud resource may result from a configuration decision rather than a software defect. In 2026, this interaction has become even more important because attackers use automation and AI to identify weaknesses faster, scale social engineering, and exploit organizational dependence on external services.

This paper therefore reframes the problem as a socio-technical risk question. Instead of asking only whether people or systems are responsible, it asks how human actions, identity controls, software vulnerabilities, third-party dependencies, and AI-enabled methods combine to produce breach conditions. The objective is to derive a current and practical prevention model that is more aligned with contemporary organizational cybersecurity.

## II. LITERATURE AND THREAT LANDSCAPE

Large-scale breach reports remain useful because they aggregate incidents across sectors and geographies. Verizon's 2025 DBIR reported a human element in roughly 60% of breaches, reinforcing the importance of credential misuse, social engineering, and mistakes [1]. However, the 2026 DBIR marks an important shift: exploitation of vulnerabilities became the leading initial access vector, accounting for 31% of breaches globally, overtaking stolen credentials for the first time in the report's history [2]. This does not make the human factor irrelevant; rather, it shows that attackers are becoming more effective at combining automated technical exploitation with human-centered intrusion techniques.

The 2026 DBIR also reports strong growth in third-party involvement and increasing risk from unapproved AI use. These findings broaden the causal model beyond employees and internal infrastructure. A breach may now originate through a supplier, a managed service, an

exposed API, or an employee entering sensitive information into an unsanctioned AI tool. The attack surface is therefore organizational and ecosystem-wide.

IBM's 2026 Cost of a Data Breach research describes an "AI tipping point." The reported global average breach cost reached USD 4.99 million, while AI-driven attacks increased by 56% compared with the previous year [3]. IBM also reports that extensive security AI and automation can materially reduce breach cost, demonstrating that AI acts as both a threat multiplier and a defensive capability. In India, IBM reported an average breach cost of INR 255 million in 2026, with phishing remaining a leading initial vector and AI-generated malicious activity becoming increasingly visible [4].

ENISA's 2026 Threat Landscape similarly emphasizes the continuing convergence of cybercrime, geopolitical activity, exploitation of digital dependencies, and emerging AI-assisted operations [5]. Taken together, these sources support a contemporary view of data breaches as multi-stage events shaped by human behavior, technical weaknesses, digital supply chains, and accelerated attacker capability.

### **III. PROBLEM STATEMENT**

Organizations continue to invest in security technologies, yet breach frequency, impact, and complexity remain high. A key reason is that security programs are often organized into separate silos: awareness teams focus on user behavior, infrastructure teams focus on patching and configuration, identity teams focus on authentication, and governance teams focus on suppliers and compliance. Attackers do not respect these boundaries. They chain weaknesses together.

The research problem is therefore not simply to identify whether human error or system vulnerability is the larger cause. The more useful problem is to determine how these factors interact, how the balance is changing under AI-accelerated attack conditions, and which integrated controls reduce the probability and impact of breaches most effectively.

### **IV. RESEARCH OBJECTIVES**

1. To review the contribution of human factors to organizational data breaches, including phishing, credential misuse, misconfiguration, and insider actions.
2. To examine the contribution of technical weaknesses, particularly vulnerability exploitation, insecure design, exposed services, weak identity controls, and delayed remediation.

3. To assess the growing role of AI-enabled attacks, shadow AI, and automated exploitation in the current threat landscape.
4. To analyse the interaction between human, technical, and third-party factors rather than treating them as independent causes.
5. To propose an integrated breach-prevention framework suitable for contemporary organizations.

## **V. RESEARCH METHODOLOGY**

This study follows a descriptive and analytical review design based on secondary data. The source base includes recent editions of major incident-analysis and breach-cost reports, supported by cybersecurity guidance and threat-landscape publications. Particular emphasis is placed on Verizon DBIR 2025 and 2026, IBM Cost of a Data Breach 2026, and ENISA Threat Landscape 2026 because these sources aggregate large volumes of observed incidents and provide current indicators on initial access, human involvement, vulnerability exploitation, third-party risk, AI-enabled threats, and breach impact.

The analysis uses thematic synthesis. Reported breach causes are grouped into five connected domains: human behavior, identity and access, software and configuration weaknesses, third-party dependencies, and AI-related risk. The method is not intended to produce a new causal percentage from primary incident records; rather, it compares reported trends and uses them to construct an updated explanatory framework.

## **VI. EVOLVING CAUSES OF DATA BREACHES**

### **A. Human and Social Factors**

Human involvement remains a major breach component. Phishing, social engineering, credential reuse, accidental disclosure, weak administrative practice, and insider misuse continue to create opportunities for attackers. The form of social engineering is also changing. Attackers increasingly use text, voice, deepfake content, and highly personalized messages rather than relying only on conventional email phishing. AI lowers the cost of producing convincing lures and increases the speed at which they can be adapted to a target. Training is necessary but insufficient. Security awareness has limited value if a successful phish immediately leads to account takeover because strong authentication, conditional access, or anomaly detection is absent. Human-centered controls therefore need to be supported by technical guardrails.

## **B. System Vulnerabilities and Configuration Weaknesses**

Unpatched software, internet-facing devices, insecure APIs, weak encryption, legacy components, and configuration errors remain major technical causes. The 2026 DBIR's finding that vulnerability exploitation became the top initial access vector is especially significant because it indicates that technical exposure is moving from a secondary factor to a primary breach entry path [2].

The operational issue is often not the existence of a patch but the delay between disclosure, remediation, and attacker exploitation. AI-assisted discovery and automation can compress this interval. Organizations therefore need asset visibility, vulnerability prioritization, rapid remediation for internet-facing systems, and compensating controls when patching cannot be immediate.

## **C. Identity and Credential Risk**

Credentials sit at the intersection of human behavior and system design. Password reuse and phishing are human-related, but the damage they cause depends on technical controls such as multifactor authentication, session management, privilege boundaries, and detection of abnormal login behavior. Identity should therefore be treated as an independent control plane rather than merely a subset of awareness training.

Privileged and service identities require particular attention because compromise can enable lateral movement, data extraction, and persistence. Modern breach prevention should reduce standing privilege, enforce phishing-resistant authentication where feasible, and continuously evaluate access context.

## **D. Third-Party and Supply-Chain Exposure**

Organizations increasingly depend on cloud providers, software vendors, contractors, managed services, and interconnected platforms. This creates breach paths that cannot be controlled solely through internal employee training or internal patching. The 2026 DBIR highlights substantial third-party involvement, showing that ecosystem risk has become a core breach driver [2].

Effective control requires supplier due diligence, contractual security requirements, segmentation, restricted third-party access, monitoring of external identities, and incident-response procedures that account for supplier-originated compromise.

## **E. AI-Enabled and Shadow-AI Risk**

AI introduces two distinct risks. First, attackers use AI to improve reconnaissance, social engineering, impersonation, malware development, and vulnerability exploitation. IBM reports a 56% increase in AI-driven attacks in its 2026 study [3]. Second, organizations create

new exposure when employees use unapproved AI services with confidential information. This “shadow AI” problem resembles earlier shadow-IT risk but can involve direct disclosure of proprietary text, source code, customer data, or internal documents.

AI governance should therefore be integrated into data-loss prevention, acceptable-use policy, model access control, logging, vendor assessment, and employee guidance. Blocking all AI use is generally unrealistic; controlled adoption with approved tools and data-handling rules is more practical.

## VII. COMPARATIVE TREND ANALYSIS

| Risk area                  | Earlier emphasis                                       | Current 2026 trend                                                                             | Priority control response                                       |
|----------------------------|--------------------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Human involvement          | Major contributor; phishing, credential misuse, errors | Still high, but increasingly combined with mobile/social-engineering and AI-assisted deception | Awareness + phishing-resistant MFA + behavioral detection       |
| Vulnerability exploitation | Growing initial-access method                          | Leading initial-access vector in 2026 DBIR                                                     | Asset inventory + rapid patching + exposure management          |
| Credentials / identity     | Leading or near-leading breach path                    | Still critical; identity compromise often amplifies other attack paths                         | Least privilege + MFA + session/identity monitoring             |
| Third-party risk           | Important but often treated separately                 | Large and growing share of breach involvement                                                  | Supplier governance + segmentation + restricted external access |
| AI-related risk            | Emerging concern                                       | AI-enabled attacks and shadow-AI leakage are material risks                                    | AI governance + secure approved tools + DLP + monitoring        |

## VIII. INTEGRATED BREACH-PREVENTION FRAMEWORK

The review supports an integrated framework built around six control layers. First, identity resilience should reduce dependence on reusable passwords and excessive privilege. Second, exposure reduction should prioritize externally reachable vulnerabilities and insecure configurations. Third, human-risk controls should combine practical awareness with technical safeguards that limit the consequences of mistakes. Fourth, third-party governance should extend access control and monitoring beyond the organizational perimeter. Fifth, AI governance should define approved tools, prohibited data classes, logging requirements, and model-security responsibilities. Sixth, continuous detection and response should connect telemetry across identity, endpoints, cloud systems, applications, and external services.

This framework reflects an important shift in cybersecurity thinking: the objective is not to eliminate human error or every software flaw, both of which are unrealistic goals. The objective is to prevent a single weakness from becoming a full breach. Layered controls should interrupt the attack chain at multiple points.

## **IX. FINDINGS AND DISCUSSION**

First, human factors remain central, but the phrase “human error” is too narrow for current breach analysis. Social engineering, identity compromise, administrative decisions, and unsafe AI usage are better viewed as components of a broader human-and-identity risk domain.

Second, system vulnerabilities have become more prominent. The 2026 DBIR indicates that vulnerability exploitation now leads global initial-access vectors, showing that patching and exposure management require the same executive attention traditionally given to phishing and credentials [2].

Third, breach causation is increasingly ecosystem-based. Third parties, SaaS platforms, cloud services, APIs, and AI tools extend the security boundary. Organizations can maintain strong internal controls and still be exposed through external dependencies.

Fourth, AI has changed both attack economics and defense. It enables faster and more scalable adversarial activity, but security automation can also reduce detection and response cost. The practical issue is therefore not whether AI is “good” or “bad” for security, but whether organizations govern and operationalize it responsibly [3].

Finally, the evidence supports a socio-technical model of breach prevention. The highest-value controls are those that reduce dependency on perfect human behavior, shorten exposure windows, constrain privilege, and limit lateral movement after an initial compromise.

## **X. CONCLUSION AND FUTURE SCOPE**

The traditional debate over whether data breaches are caused mainly by people or by technology is increasingly inadequate. Current evidence shows that organizational breaches emerge from interconnected weaknesses involving human behavior, identity, software vulnerabilities, configuration choices, suppliers, cloud platforms, and AI-enabled methods. Human involvement remains substantial, yet vulnerability exploitation has become a leading entry path, while AI and third-party dependencies are reshaping both attack speed and breach impact.

Organizations should therefore move away from one-dimensional security programs. Effective breach prevention requires coordinated identity protection, rapid vulnerability remediation, secure configuration, supplier controls, AI governance, continuous monitoring, and practical user education. Future research should test this integrated framework using organization-level incident datasets, sector-specific surveys, and longitudinal measurements of control effectiveness. Additional work is also needed on AI-model compromise, agentic AI risk, deepfake-enabled fraud, and metrics for measuring the security value of AI automation.

## REFERENCES

1. Verizon Business, “2025 Data Breach Investigations Report,” Verizon, 2025.
2. Verizon Business, “2026 Data Breach Investigations Report,” Verizon, 2026.
3. IBM and Ponemon Institute, “Cost of a Data Breach Report 2026,” IBM, 2026.
4. IBM India, “India Records its Highest Average Cost of a Data Breach at INR 255 Million in 2026,” Aug. 3, 2026.
5. European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2026,” Sep. 22, 2026.
6. European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2025,” Oct. 1, 2025.
7. National Institute of Standards and Technology, “The NIST Cybersecurity Framework (CSF) 2.0,” NIST CSWP 29, 2024.
8. Cybersecurity and Infrastructure Security Agency, “Secure by Design,” CISA, current guidance.
9. OWASP Foundation, “OWASP Top 10: Web Application Security Risks,” OWASP, current edition.
10. Identity Theft Resource Center, “Annual Data Breach Report 2025,” ITRC, 2025.