BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

An Explainable Risk-Aware Framework for Governing and Monitoring Shadow AI in Enterprise Environments

Shradha Vilas Kharabi

Student, Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India

Prof. Supriya Survase

Guide, Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India

Abstract

Generative AI assistants are now being used for work purposes, often beyond the view of enterprise IT, security, and compliance functions, creating a shadowy14 area where data usage is difficult to track and responsibility hard to establish. By analyzing six different papers mentioned in the seminar synopsis together with related standards and explainability literature, this study identifies gaps in research on detection, classification of risk, and governance response. It proposes a conceptual six-step framework (Monitoring and Discovery, Context Collection, Risk Assessment, Explainability, Governance Decision, and Continuous Monitoring) which aims to turn authorized monitoring signals into context-enriched events scored using a transparent if formula with hard-rule floors and an uncertainty band, fully traceable reason codes, and proportionate response options including a specific "indeterminate" category. The study is based on conceptual design informed by a reference-driven literature review: no algorithms, datasets, or experiments have been implemented, and all weights, thresholds, and scenarios are purely illustrative. An evaluation plan is suggested, covering thematic areas of detection adequacy, classification coherence, explanation value, latency, privacy, and ablation analysis. The framework is intended to provide decision-makers with the needed context for responsible Shadow AI governance; its effectiveness still needs to be proven in practice.

Keywords : Shadow AI, Enterprise Cybersecurity, AI Governance, Explainable Artificial Intelligence, Risk Assessment, Security Monitoring, Policy Enforcement.

I. Introduction

A. Background and Motivation

Generative AI assistants such as chatgpt, Gemini, Claude. Github Copilot have rapidly flooded the enterprise with less control than many organizations have time to sanction, configure, and govern. In this paper, Shadow AI is the use of AI tools or services not covered by an organization's processes or

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

sufficiently governed. Extending the Shadow IT phenomenon -- in which employees adopt technology not sanctioned by IT [1], [2] --into a domain where the tool receives free-form text, code, or documents and responds with generated output [3], [4], reviewed sources describe Shadow AI as a governance issue that incorporates unmanaged data flows, visibility. Accountability [3]–[5].

Enterprise security, privacy, and compliance teams therefore face a practical question: how can AI-tool use be monitored, evaluated. Governed in a manner that is proportionate, understandable to the people who must act on it, and compatible with business continuity? Prohibition is blunt, and the reviewed literature favors controlled enablement with registries, role-aware policy, audits, and escalation [4], [5].

B. Problem Statement

Starting from the problem framing in the underlying seminar synopsis, we state the problem as the conjunction of the following concerns: (i) limited organizational visibility into which AI tools are used; (ii) accidental exposure of sensitive or confidential information; (iii) cybersecurity and (iv) privacy risks; (v) regulatory and organizational compliance obligations; (vi) inadequate accountability; (vii) difficulty detecting and classifying risky activity; (viii) the absence of transparent explanations for risk decisions; and (ix) fragmented monitoring, risk assessment. Governance mechanisms. Items (i)–(vi) are supported by the reviewed sources [3]–[6]; items (vii)–(ix) are the synopsis' analytical observation about the current state of the field and are examined critically in Section III.

C. Research Objectives and Questions

The work pursues four objectives: O1, to investigate Shadow AI detection approaches; O2, to analyze Shadow AI security, privacy. Compliance risks; O3, to examine Explainable AI (XAI) for risk assessment; and O4, to propose an explainable, risk-aware governance framework. They are operationalized as four research questions:

• RQ1 (O1). Which authorized signals can support discovery of AI-tool usage, and what visibility limits constrain them?

• RQ2 (O2). Which risk dimensions are needed to express security, privacy, and compliance exposure of an individual AI-usage event?

• RQ3 (O3). How can a risk decision be made traceable and understandable to security, IT, and compliance reviewers without relying on opaque models?

• RQ4 (O4). How can risk outcomes, including uncertain ones, be mapped to proportionate governance actions and fed back into policy refinement?

D. Contributions

This paper makes five contributions, all conceptual: (C1) a critical, source-qualified comparison of six sources named in the synopsis, with explicit disclosure of what could and could not be verified; (C2) a refinement of the five research gaps stated in the synopsis, each tied to evidence and to a design response (Table II); (C3) a six-module architecture with defined inputs, processing, outputs. Data flow (Section V); (C4) an explainable risk model that combines a normalized weighted score, hard-rule floors, and an uncertainty interval that yields an explicit "indeterminate" outcome (Section VI); and (C5) a threat analysis, illustrative scenarios, and a validation plan that separate proposed design from demonstrated

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

results (Sections VII–IX). The paper does not claim implementation, empirical results, or priority over existing work.

The remainder of the paper is organized as follows. Section II reviews related work; Section III defines the problem and gaps; Section IV describes the methodology; Section V presents the architecture; Section VI details the risk and explanation model; Sections VII–IX cover threats, scenarios. Evaluation; Sections X–XII discuss, list limitations, and conclude.II.

II. Background and Related Work

A.Shadow AI and corporate security

Puthal and colleagues provide a survey-level analysis in which Shadow AI is presented as an extension of the current cyber-threat environment, with a focus on the unauthorized handling of sensitive data, data leakage, and threats related to models, and they also discuss possible ways to address these issues [3]. The examples they give are mainly conceptual and not based on data from actual organizations. Silic and co-authors apply the Shadow IT concept to the field of artificial intelligence; the publisher's abstract refers to a mixed-methods study that combines a survey of practitioners with interviews of executives and suggests the establishment of registries, training programmes, audits, and escalation procedures [4]. For this source, only information at the abstract level was available. A second conference paper led by Puthal and a paper on governance and resilience by Ross and others are listed under similar titles in the synopsis [7], [8]; their bibliographic details could be verified only via secondary citations, and so no specific findings are assigned to them here other than those relating to the topics they cover.

B. The detection and monitoring of Shadow AI

Two contributions provide specific designs for detection or classification. Anuguthala suggests a three-layer method made up of a use-case registry, enforcement of CI/CD practices, and a genai gateway, with the aim of addressing AI workloads involved in software development and providing access to genai services [6]. However, the evaluation described is small and partly based on synthetic data, and certain baseline detection results are taken from previous literature rather than being obtained within the same experiment, which hampers the ability to make comparative claims; furthermore, the study is authored by a single person and focuses on engineering workloads rather than on interactive chat use [6]. Ojowa et al. Propose a detect–classify–govern pipeline incorporating multi-signal detection, a multi-dimensional risk taxonomy, a staged governance lifecycle, and remediation categories (block, replace, regularize), but present it as a framework that has not yet been validated and have planned validation [5]. Taken together, they indicate that registry-based identification and the use of layered signals could form a reasonable basis, while leaving undecided the question of how interactive usage events should be scored and explained.

C. Enterprise AI Governance

The governance approach in the set of cases highlighted emphasizes accountability, the careful provision of enablement rather than prohibition, and the existence of escalation routes [4], [5]. The literature on Shadow IT also confirms this point: unmanaged adoption of technology is generally a sign that certain business needs have not been satisfied, and therefore strict punitive policies will have the effect of displacing rather than eliminating it [1], [2]. It is for these reasons that the proposed design calls for the standard reaction to unauthorised use to be one of redirection and review, not automatic denial.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

D. Explainable Risk Assessment

The literature on XAI separates inherently interpretable models from the use of post-hoc explanations for opaque models [9], [10]. Rudin states that when it comes to important decisions an interpretable model should be chosen over explaining a black box [10]. Techniques that are post-hoc, such as LIME and SHAP, provide approximations of local feature attributions for complex models [11], [12]. According to Miller's review of the social sciences, useful explanations need to be contrastive, selective, and socially situated [13], and research into explanation metrics stresses that the quality of an explanation must be assessed by users and not assumed [14]. Nothing among the six synopsis sources, as far as could be verified, assesses the quality of explanations for decisions involving Shadow AI risk.

E. Considerations regarding standards and regulation

The NIST AI Risk Management Framework categorises AI risk management activities under five functions: Govern, Map, Measure, and Manage [15], whereas the NIST Cybersecurity Framework 2.0 divides cybersecurity outcomes into six parts: Govern, Identify, Protect, Detect, Respond, and Recover [16]. ISO/IEC 42001 provides a framework for an AI management system [17]. NIST SP 800-30 describes a risk assessment procedure which includes judgments about threat, vulnerability, likelihood, and impact [18], and Zero Trust Architecture provides for the separation of policy decision and enforcement [19]. In the European Union, the AI Act [20] and the GDPR [21] lay down the obligations that apply to AI systems and personal data. The current paper uses these documents as design references and does not claim conformity with them nor give any legal advice.

The comparison of the sources is given in Table I, and in the 'evidence status' column the authors of the paper indicate those items which they have found to be verifiable, without making any kind of quality assessment.

Table I: Comparison of the six synopsis sources (verified status reflects access achieved)

Source	**Type / method (as verified)**	**Contribution relevant here**	**Evidence status**	**Limitation for this work**
Anuguthala [6]	Industry-affiliated experimental study; registry, CI/CD gate, genai gateway; full text read	Registry-aware prevention of AI workloads; policy enforcement at build and access time	Small evaluation, partly synthetic; some comparators taken from literature	Engineering workloads, not interactive chat use; large-enterprise setting
Ojowa *et al.* [5]	Preprint (not peer reviewed); literature-informed framework; full text read	Detect–classify–govern pipeline; risk taxonomy; governance lifecycle; block/replace/regularize	Validation planned, not executed	No empirical results; scoring and explanation not specified in depth

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Puthal *et al.* [3]	Peer-reviewed survey-style article; full text read	Threat landscape, data-leakage and security implications, mitigation directions	Conceptual; hypothetical illustrations	No organizational measurement; no integrated pipeline
Silic *et al.* [4]	Peer-reviewed; mixed methods per publisher abstract; abstract only	Shadow IT to Shadow AI extension; registries, training, audits, escalation	Survey plus interviews (per abstract)	Strategies not validated for effectiveness (per abstract); detail unverified
Puthal *et al.* (OCIT) [7]	Conference paper; record via secondary citation; not accessed	Topic: attack surface, risks, countermeasures	Unverified	Findings cannot be assessed
Ross *et al.* [8]	Conference paper; record via secondary citation; not accessed	Topic: governance, risk, organizational resilience	Unverified	Findings cannot be assessed

III. Problem Definition and Research Gaps

Let an AI-usage event e denote a single observed interaction between an enterprise subject and an AI service. The governance problem consists in assigning to every e a reviewable risk assessment and a proportionate action, relying solely on authorized signals, under incomplete context, with an explanation adequate for an accountable human to accept, contest, or override the outcome.

The synopsis identifies five gaps. We regard these as hypotheses concerning the literature rather than as established universal facts, and we qualify them against the verified evidence (Table II). Two qualifications are significant. First, the registry-aware design in [6] does indeed address build-time and runtime prevention for AI workloads; consequently, Gap 4 is best understood as incompletely addressed for interactive end-user usage and for explained risk decisions, rather than as unexplored. Second, [4] presents empirical work; therefore, Gap 1 pertains to integrated governance frameworks rather than to all Shadow AI research. Since two sources could not be assessed, absence claims are confined to the verified subset.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Table II : Research gaps, evidence, proposed design response, and remaining validation need

Gap (synopsis)	Supporting literature	Limitation of existing approaches	Proposed response	Remaining question
G1: Frameworks conceptual or without completed validation	[5]: validation planned; [3]: survey-level	Cannot show that detection, scoring, or actions work	Specified model and testable metrics (Section IX)	Does an implementation meet targets on realistic data?
G2: Small samples; smaller organizations under-covered	[6]: three practitioner participants, large enterprises; [4]: sample coverage unverified	Limited generalizability to smes	Modular, tiered deployment; minimal-signal profile for small organizations	Which modules are viable with limited telemetry?
G3: Detection, classification, enforcement not integrated and validated	[5]: integrated but unvalidated; [6]: integrated for workloads	No verified end-to-end approach covering interactive use with explanation	Six-module pipeline with explicit data contracts (Section V)	Does integration outperform component baselines?
G4: Reactive detection dominates; prevention partial	[6]: build-time and gateway controls; others reactive or conceptual	Prevention evidence limited to workloads	Risk tiers drive pre-emptive redirection and conditional restriction	Can prevention for interactive use avoid excess friction?
G5: Little benchmarking against each other or regulation	[5]: conceptual mapping; [15], [20]	Weak cross-study comparability	Indicative alignment to NIST functions; shared metrics and baselines	How does the design compare quantitatively and in audit terms?

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

IV. Research Methodology and Framework Design Approach

The study constitutes conceptual framework development informed by a reference-driven literature review and design analysis, loosely following the problem-identification, objectives, design, demonstration. Evaluation-planning stages of design science research [22]. The stages actually carried out were: (1) problem and objective extraction from the seminar synopsis; (2) location of each of the six synopsis references by title search, followed by reading of the accessible record; (3) addition of standards and explainability sources selected for relevance; (4) refinement of the gaps; (5) design of the six modules and risk model; (6) construction of illustrative. Scenarios as an analytical walkthrough; and (7) drafting of an evaluation plan.

This was not a systematic review. No database protocol, screening counts, or PRISMA flow were generated, and the selection of additional sources is purposive and may be subject to bias. Full text was read for [6], [5]. [3]; only the publisher record or abstract was available for [4]; and [7] and [8] could be confirmed solely through secondary citations. Several bibliographic details in the synopsis (titles, author initials, venue names) differed from the records located and were corrected in this paper; the corrections are documented in the accompanying verification report. The design stage represents the authors' own analysis. Each design decision is linked to a gap, and every numerical parameter is flagged as illustrative

V. Proposed Framework Architecture

A. Overview and Data Flow

Figure 1 shows the framework as a pipeline from enterprise AI-usage sources down to enterprise controls, and back up through a Continuous Monitoring component that closes the loop by returning the reviewed outcome as governed, versioned changes. This pipeline maintains a single evolving event record: the Monitoring and Discovery layer generates it, Context Collection appends attributes and provenance, Risk Assessment appends scores and a tier, Explainability appends rationales. Governance Decision appends an action and accountability. Table III summarizes the contracts for each module. Each module is a proposed design element, none has been implemented.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

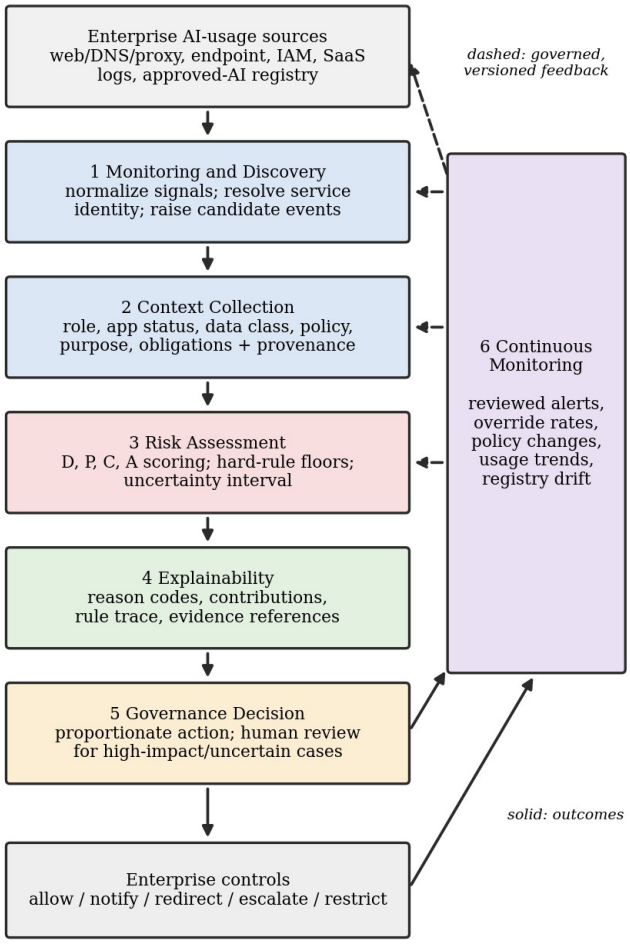


Fig. 1. Overall proposed framework. Solid arrows show event flow; dashed arrows show governed, versioned feedback from Continuous Monitoring.

Table III : Module inputs, processing, and outputs

Module	**Inputs**	**Processing**	**Outputs**	**Consumed by**
1 Monitoring and Discovery	Proxy/DNS/gateway logs, endpoint and IAM events, saas audit logs, approved-AI registry	Normalization, service-identity resolution (aliases, domains, apis), confidence tagging	Candidate event with service ID and discovery confidence	Context Collection

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

2 Context Collection	Candidate event, directory, asset and data-classification stores, policy repository	Attribute lookup, state labeling (known/inferred/unknown/conflicting), provenance, minimization	Context-enriched event	Risk Assessment
3 Risk Assessment	Enriched event, scoring configuration, rule floors	Component scores D, P, C, A; interval bounds; floors; tiering	Score, interval, tier or indeterminate	Explainability
4 Explainability	Scored event, rule trace, evidence references	Reason codes, contribution summary, contrastive hint, audience views	Explanation record	Governance Decision; reviewers
5 Governance Decision	Explanation record, action policy, control inventory, modifiers	Tier-to-action mapping, modifiers, human-review routing	Action, rationale, owner, review flag	Enterprise controls; Continuous Monitoring
6 Continuous Monitoring	Decisions, reviewer outcomes, policy changes, inventory drift	Aggregation, override analysis, change proposals	Versioned policy or threshold change proposals	Modules 1, 2, 3, 5 via change control

B. Monitoring and Discovery Module

This layer detects use of AI tools originating from authorized sources. Candidate sources under consideration as inputs to this layer include secure web gateway and proxy logs, DNS logs, network telemetry, endpoint security events, identity and access-management events (e.g., oauth consent grants), saas audit logs, enterprise AI application catalogs and approved-AI registry. These are candidate design inputs; this layer does not yet assert these as its capabilities. Layered detection conforms with the multi-signal approach of [5], and discovery through registry lookup is the basis for the gateway design of [6]. Matching against domains, endpoints and known aliases in a catalog of known service identifiers resolves each candidate event to a service identity, generating a degree of confidence alongside the event.

It is necessary to explicitly treat the limitations of visibility as constraints. Since encrypted traffic can only be subject to metadata analysis unless inspection is specifically authorized, the end points might not be

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

protected or could be personal ones; services might hide themselves or might include AI within another application; and the traffic could go through private cellular or other networks which remain invisible to the enterprise. For this reason, the module has to report the extent of its discovery by type of signal in order to show when there is a complete lack of evidence. This in turn enables further reasoning about the possibility of false negatives.

C. Context Collection Module

Context is important in determining whether a particular use is innocent or harmful. For each event, the Context Collection module captures at least the following: the user's role and permissions, the identity and approval status of the AI application, the sensitivity of the data, the relevant organizational policy, the intended business purpose (if this has been specified), the applicable obligations, and the properties of the destination. For each attribute, the collection describes its status (known, inferred, unknown or conflicting), origin and timestamps. Conflicts (e.g. A registry indicating a service as approved while a policy repository denies it for a specific data category) are explicitly identified and resolved assuming a declared priority of source. If any relevant attributes are absent or given incorrectly, the event takes the uncertainty branch defined in Section VI.

Since this layer handles employee data, minimization is a design objective: processing content is preferable to inspecting payload, using pseudonymous identifiers (re-identification permitted only for authorized research), restricting access by role, enforcing retention schedules. Applying purpose limitations, including a data protection impact assessment where warranted [21]. Whether content inspection should be conducted is an operational and legal question, beyond the scope of this work.

D. Risk Assessment Module

Within this module, the calculation of the component scores and tier of Fig. 3 is housed, and hence requires only the context record plus a versioned configuration in order to perform its operations. As all results are derived from the context record, they are auditable and can be reconstructed from the records held in the database.

E. Explainability Module

This module takes the output and rule trace of Risk Assessment and transforms them into an explanation record as per Section VI-D, but adapted to the audience: a one-sentence justification for the affected user, factor and rule view for the IT/security analyst. An obligation and evidence view for the legal/compliance reviewer.

F. Governance Decision Module

This module transforms the tier into decisions regarding response, taking into account business-criticality, role, recidivism, presence of an approved, and availability of an immediate remedy. Escalates matters to human reviewers, as per Section VI-E, and keeps a record of the decision, including an accountable owner and rationale.

G. Continuous Monitoring Module

This module stores decisions, reviews, policy changes, and inventory changes, including the three distinct flavors of change: modifications to policies (e.g., addition of an approved service), recalibration of

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

thresholds/weights. Updates to ML models (not part of the baseline design, but presumably addressed in a separate ML lifecycle). Nothing in the baseline design updates itself or requires retraining. Monitors useful metrics as described in Section VI-G, including overrides by reviewers, precision after review, time to triage, newly discovered services, and usage patterns by role. The correspondence to established frameworks like NIST CSF, AI RMF [15], [16], etc. Is illustrative: the discovery and context modules correspond to NIST CSF Identify, Detect. The AI RMF Map; scoring corresponds to AI RMF Measure; decision-making corresponds to AI RMF Manage and CSF Respond; and Continuous Monitoring corresponds to the Govern function.

VI. Risk Assessment and Explainable Governance Model

A. Risk Dimensions

For an observed event e, four component scores are defined, which take values in [0,1] (Table IV). Each is based on ordinal judgments that are mapped to the interval, so that any final score is a prioritization rather than a probability of harm. The use of formulaic, documented scoring is not novel; it has been advanced elsewhere in the literature, for instance, by CVSS [23]. The need to combine judgments about impact and likelihood is noted in [18]. The current candidate model falls within these precedents as a calibration target.

Table IV : risk dimensions and illustrative scoring anchors

Dimension	Meaning	Example indicators	Illustrative anchors
D(e): data sensitivity	Sensitivity of data likely or known to be involved	Data class of source system, labels, pattern matches, repository class	Public 0; internal 0.33; confidential 0.67; regulated or restricted 1.0
P(e): policy violation	Degree to which the event departs from applicable policy	Service status in registry; permitted data classes; role entitlements	Approved and permitted 0; approved but out of scope 0.5–0.8; unapproved 1.0
C(e): compliance exposure	Additional obligations triggered beyond data class itself	Personal-data transfer, absent processing agreement, sector or contractual rules	None 0; possible 0.3; probable 0.6; definite 0.8–1.0
A(e): service trust	Assurance level of the destination service	Contract status, retention and training terms, security attestations, provenance	Enterprise-contracted 0.1; reviewed consumer 0.5; unknown consumer 0.9

D(e) captures the information sensitivity, and C(e) is about the incremental obligations (e.g., a cross border transfer, missing agreement) not the information class. To avoid double counting However, there are

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

intrinsic correlations, in particular between D and C and between P and A . This is addressed in the ablation and weight-sensitivity experiments in Section IX .

B. Scoring Model:

The candidate score is a weighted sum form, scaled to [0,100]:

R(e) = 100.(w_D*D(e)+w_P*P(e)+w_C*C(e)+w_A*A(e)) (1)

The four weights are non-negative and sum to 1. For example, the weights are selected as wd = 0.35, wp = 0.20, wc = 0.25, and wa = 0.20. These weights are not optimized or validated and should be elicited and validated by stakeholders in practice. Then, for illustration purposes, tiers are Low (0-25), Moderate (25-50), High (50-75), and Critical (75-100) – also unvalidated.

C. Uncertainty and Hard Floors

Rlo(e)=100*sum_k in K (w_k*S_k) Rhi(e)=100*sum_k in K (w_k*S_k(e))+sum_k in U (w_k) (3)

If R_lo and R_hi are in the same tier the event is assigned that tier with confidence, otherwise it is undetermined. Similarly the minimum and maximum values for the different options are tracked for each conflicting attribute. The ranges indicate that the utility function is compensatory, i.e., low values of one component can compensate for high values of another. But the model also encodes hard-rule floors: a rule such as “regulated data to an unapproved service implies at least High.

Algorithm 1 : Event assessment with uncertainty handling (pseudocode)

Input: event e with context; weights w; tier bounds; floor rules: F

1 for each component S in {D,P,C,A}: read state + value

2 R_lo = 100 sum over k w_k S_k (known); R_hi = R_lo + 100 sum over k w_k (unknown)

3 tier_lo = tier of R_lo; tier_hi = tier of R_hi

4 floor = max tier among applicable rules in F (record rule ids)

5 if tier_lo == tier_hi: use tier = max(tier_lo, floor)

Else if floor >= tier_hi: use tier = floor

6 else: tier = INDETERMINATE

7 build justification (Section VI-D); route according to Section VI-E

8 log inputs, versions, rule ids, tier, action

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

D. Explanation Generation

Since the base model has an explicit rule-and-weight form, explanations can be generated as precise statements. Specifically, for each event e, the Explainability module provides (a) rule-based reason codes (such as DATA_REGULATED, SERVICE_UNAPPROVED); (b) a contribution breakdown ck = 100 w k S k , for transparency. That a reviewer can evaluate what raised the score; (c) a human-readable sentence; (d) traceability to policies and floor rules, by identifiers; (e) references to attributes and their sources; and (f) a contrastive utterance of the type “the tier would be Moderate if...”, in concordance with findings that explanations benefit from being contrastive and selective [13]. Approximate explanations such as those generated by SHAP [11] and LIME [12] are not useful for this use case, since they attempt to approximate opaque models. An interpretable model is preferable, if not mandatory, whenever one is available [10]. If ML components such as content classifiers are added, post-hoc explainers would then become relevant and subject to the same evaluation [9], [14]. Fig. 2 details the process, including the indeterminate case.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

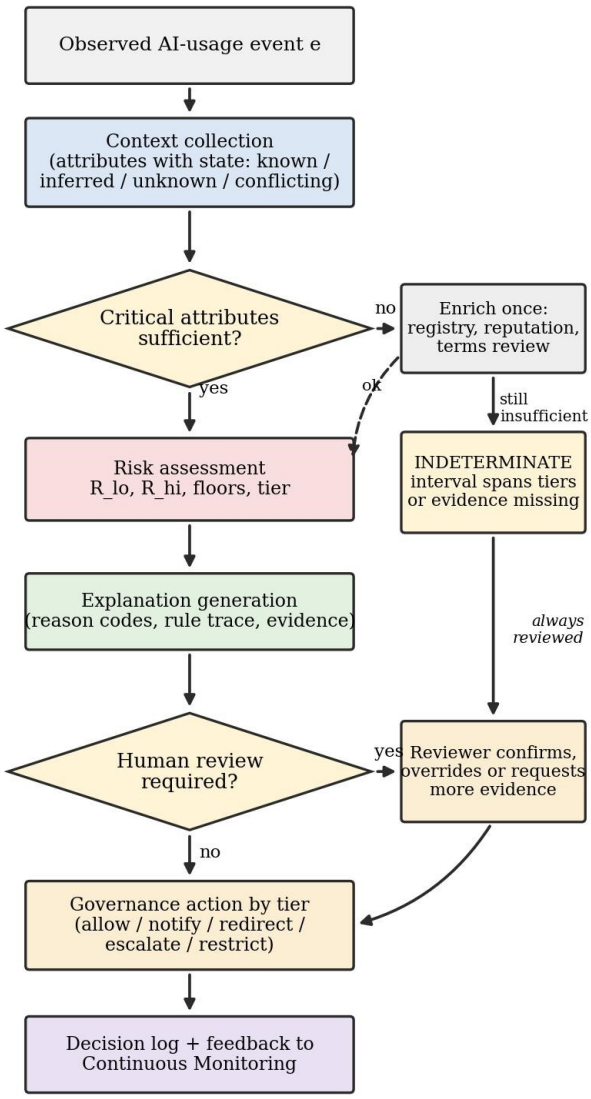


Fig. 2. Risk assessment and governance workflow. Insufficient evidence leads to an indeterminate outcome that is always reviewed by a human.

Only if a control is available, the policy allows it, and, for High and undetermined events, human confirmation is attained. Modifiers provide flexibility without undermining the tier system: for instance, a business-critical process may require human approval before any blocking occurs. However, they do not permit downgrading the tier. False positives are addressed through the process modifier, and the mapping from events to actions also includes a channel for the affected employee to dispute the outcome. Logging of the owner and the reasoning is another prerequisite for accountability, in concordance with human-in-the-loop governance in AI [15], [20]. While the model itself does not make decisions, it supports several levels of human judgment. The adequacy of this chain for the given use case, and the allocation of responsibility for false positives and negatives, must be determined for the target deployment.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

E. Governance Mapping and Human Oversight

The mapping from tiers to actions is detailed in Fig. 3. The options include allowing an approved activity under policy, continuing monitoring, requesting clarifications, recommending an approved enterprise AI service, escalating for a human-in-the-loop security or compliance review; and restricting or blocking when justified by policies and available controls.. An unapproved service is not assumed to be malicious: a Moderate tier typically maps to a notification and a recommended alternative, rather than sanctions. Similarly, High and Critical events are restricted or blocked.

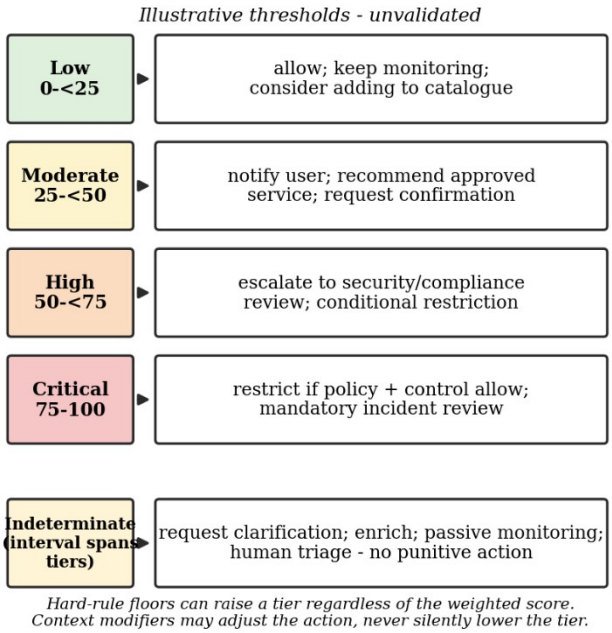


Fig. 3. Risk classification and governance mapping. Thresholds and actions are illustrative and unvalidated.

VII. Security, Privacy, and Threat Analysis

The threats and their impact are summarized in Table V. Notably, the design does not claim to prevent all harms: the model can only mitigate those for which the features are available. Its preventable harms are a subset of all possible ones. The analysis is qualitative and design-stage: no threat has been formally validated.

Table V : Threat analysis (design-level, untested)

Threat	**Effect**	**Design response and remaining concern**
Confidential-data exposure to AI services	Mitigate / identify	Data-class context and floors drive redirection or restriction; exposure that occurs outside visible channels is not addressed

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Unauthorized AI services	Mitigate / identify	Registry comparison and approval status; new services surface through discovery and are reviewed
Incomplete application identification	Identify	Discovery confidence and alias catalogue; unresolved services follow the indeterminate path
Limited monitoring visibility	Identify	Per-source coverage reporting; unmonitored devices and networks remain out of reach
Attempts to bypass controls	Identify	Anomaly review of unusual routes and shadow tunnels; determined evasion not prevented
Incorrect risk classification	Mitigate	Reason codes, rule trace, review, and override-rate monitoring enable detection and correction of errors
False positives and false negatives	Mitigate / identify	Tiered actions limit harm of false positives; false negatives from invisible usage persist and require measurement
Over-collection of employee information	Mitigate	Minimization, pseudonymization, retention limits, and access control; adherence depends on implementation and governance
Misuse of monitoring logs	Mitigate	Role-based access, audit of log access, purpose limitation; insider abuse by administrators remains a residual risk
Excessive restriction harming work	Mitigate	Proportionate mapping, approved-alternative recommendation, human confirmation before restriction
Third-party AI service risks	Identify	Trust dimension A(e) encodes known contract terms; vendor-side breaches and training-use practices are outside enterprise control

Two aspects are of particular importance. First, the framework itself is a high-value target: it stores sensitive evidence, and its integrity must be protected as security infrastructure. Defense-in-depth and zero-trust design principles apply to the framework itself and its features [19]. Second, the framework cannot make absolute safety guarantees: its prevention claims are limited by the available signals, and explicit statements of prevention without evidence are misleading.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

VIII. Illustrative Enterprise Scenarios

In the first scenario, regarding regulated customer data being used by an unauthorised public assistant, it was noted that a proxy log indicated a large upload of a paste type from a finance employee's managed laptop to a consumer chat service that is not included in the registry; the record-system data classification is regulated personal data.

Table VI : Illustrative scenario scores (hypothetical; illustrative weights)

Scenario	D	P	C	A	R or interval	Tier and action
1 Customer data to unapproved public assistant	1.0	1.0	0.8	0.9	93.0	Critical; restrict if control exists, mandatory review
2 Approved enterprise AI, permitted task	0.33	0.0	0.1	0.1	16.1	Low; allow, keep monitoring
3 AI coding assistant with restricted source	0.67	0.8	0.3	0.4	55.0	High; escalate, conditional restriction
4 Unfamiliar service, little evidence	?	1.0	?	?	[20, 100]	Indeterminate; clarify, enrich, monitor

Scenario 1:The employee may access the data but is not permitted to transfer it outside the organisation; there is no processing agreement and the retention periods are unknown. The values of the factors are D = 1.0, P = 1.0, C = 0.8 and A = 0.9, so that R = 100 × (0.35 + 0.20 + 0.20 + 0.18) = 93.0, and this is categorised as Critical, a classification which is confirmed by the floor rule. Explanation (example): "Critical because regulated personal data (DATA_REGULATED) was sent to an unapproved service (SERVICE_UNAPPROVED); rule FLOOR-01 applies; the tier would have been at most High if the destination had been an approved service." The action that is required is to restrict the activity if a control is available and the policy allows this; the user should be notified, the approved assistant should be recommended and a mandatory incident review should be opened. Since it is not possible to determine the actual content of the upload because of the encryption of the traffic, the data class is inferred from the source system rather than being observed.

Scenario 2: An enterprise AI which has already received approval is being used for a purpose that is allowed. Situation — an employee asks the enterprise-contracted assistant about internal, non-sensitive information. The values are D = 0.33, P = 0, C = 0.1, A = 0.1, so the risk rating is 100 × (0.1155 + 0 + 0.025 + 0.02) = 16.05, or approximately 16.1, which is classified as Low. What to do: permit the request and keep the minimal metadata. Advantage: this shows that the framework should not produce alerts when the AI is being used in a permitted way. Note: the content class is inferred and, in accordance with the governance rules, some sampling may be required for control validation.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

.**Scenario 3:** A developer is using an AI coding assistant that has restricted source code access. The assistant in question is one that has been approved and is licensed only for use in non-restricted repositories, yet it is being used in a repository that has been marked as confidential. The relevant values are D at 0.67, P at 0.8 (since the approved service is outside the scope), C at 0.3 (because of possible contractual restrictions), A at 0.4, and R at 100 times (0.2345 plus 0.16 plus 0.075 plus 0.08), which is 54.95 or about 55.0, a high value. The appropriate course of action is to raise the issue with both the repository owner and the security reviewer, to consider imposing a conditional restriction until the issue has been reviewed, and to recommend a permitted configuration. The uncertainty involved is whether the code was actually transmitted and whether the vendor's retention policy applies; this scenario illustrates the CI/CD-stage governance focus referred to in [6].

Scenario 4: An AI service that is unfamiliar and for which there is not sufficient evidence. Event: The DNS and proxy logs show repeated access to an unknown domain linked to AI, the request sizes being small. Context: P equals 1.0 (it is not in the registry); D, C, and A are unknown. The interval is Rlo = 100 × 0.20 = 20 and Rhi = 100 × 1.0 = 100, so all the tiers within the range are included. Classification: indeterminate. Explanation: "The risk cannot be determined: the service is not registered, but the data class, the obligations, and the service trust are unknown." Action: Carry out one enrichment attempt (by checking the registry, examining the reputation, and reviewing the terms), ask the user for clarification using non-punitive language, continue with passive monitoring, and have a human perform triage; no blocking is carried out. Uncertainty: stays in place until further evidence is obtained; this scenario demonstrates the advantage of having an explicit unknown state rather than forcing a score.

IX. Method of Evaluation and Plan of Validation

No implementation has been carried out and no experiments have been conducted. The account given below relates to a controlled evaluation which has still to be carried out and contains no numerical results, whether reported or implied.

A. Environment, Data, and Baselines

A prototype would be used in a laboratory environment that is similar to an actual enterprise one (it includes managed endpoints, logging through a proxy and DNS, a mock approved-AI registry, a labelled data-classification store, and a policy repository). The test events would be synthetic or generated using scenario templates, the true labels being decided by an expert panel on the basis of written criteria, and they would include benign, borderline, risky, and ambiguous cases as well as event streams with injected missing and conflicting attributes. Actual employee data would not be employed; any pilot carried out within an organisation would need to go through an ethics and legal review. The baseline approaches are: B0, the allow/deny lists in the registry; B1, keyword or pattern-based DLP-style rules; B2, binary policy rules without scoring; and B3, the proposed scoring without explanations or the uncertainty interval.

B. Metrics

The metrics are presented in Table VII along with their relationship to the objectives and the research questions and the kind of analysis planned; for the main comparisons, confidence intervals derived from bootstrap resampling and paired tests between the baselines will be used and effect sizes, not significance alone, will be reported. The sensitivity of the weight and threshold will be investigated by modifying the parameters and then noting the resulting changes in the tier assignments.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Table VII : proposed evaluation metrics (planned; no results exist)

Metric	How measured	Objective / RQ	Comparison or analysis
Detection precision, recall; FP and FN rates	Against labeled event streams, per service and per tier	O1 / RQ1	B0, B1, B2; per-source ablation
Risk-classification consistency	Agreement (e.g., Cohen's kappa) with expert panel; repeatability on identical inputs	O2 / RQ2	B2, B3; weight perturbation
Explanation completeness and usefulness	Checklist coverage of reason codes and evidence; reviewer study measuring decision accuracy, time, and ratings	O3 / RQ3	B3 vs full; guided by [14]
Policy-to-action correctness	Fraction of events mapped to the action prescribed by policy; trace audit	O4 / RQ4	Manual policy walk-through
Latency and resource overhead	Per-event processing time; CPU and memory under load	O4	Event-rate scaling
Scalability	Throughput as events per second and number of subjects increase	O4	Stress test
Robustness to missing context	Performance when attributes are removed at 10–50% rates	O2, O4 / RQ2	Interval vs imputation
Privacy and minimization	Field-level audit of stored data; re-identification attempt on logs; retention checks	O4	Minimal vs full collection
Human-review effort	Alerts per reviewer-hour; share of events routed to review	O4 / RQ4	B0–B3
Ablation	Remove each of: context module, floors, interval, explanations	O3, O4	Contribution of each element

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Factors which might affect the validity of the study include the realism of the synthetic data, the subjectivity of the experts' labels, and the small number of reviewers; these points will be provided together with the results.

X. Discussion

The suggestion is to integrate the system with a clear approach to dealing with uncertainty. The methods mentioned in [6] and [5] show that detection and governance can be linked; the present design features a transparent event-level risk model whose results are by nature explainable, as well as a definite procedure for handling insufficient evidence. While a tiered mapping may reduce unnecessary blocking when compared to a simple allow/deny registry, this is still only a hypothesis that needs to be tested and not a conclusion.

The different design trade-offs are evident. Although transparent linear scoring is easy to audit it is less expressive than machine learning classifiers; hard-rule floors help to address compensation problems but do so at the cost of greater maintenance requirements for the rules; the interval method removes false precision but might increase the workload on humans. In an operational setting the framework needs current registries and data classification, a requirement which smaller organisations may not be able to meet; a basic profile that makes use only of registry, proxy, and role information could serve as a practical starting point, and it is still an open question as to whether this is sufficient (Gap G2). Adoption also depends on trust: if the monitoring is seen by employees as intrusive it can harm cooperation, so it is necessary to include in the governance design details about what is collected and why.

XI. Limitation And Future Work

The work is grounded in conceptual ideas because it includes no actual implementation, dataset, or experiment, the parameters provided being only illustrative. The degree to which things can be observed is restricted by factors such as encryption and unmanaged devices. Of the six sources mentioned in the synopsis, only two could be verified by reference to other sources and of these only one could be confirmed at the abstract level; as a result the synthesis of the literature is to some extent limited and the additional sources were deliberately chosen rather than selected in a systematic way. The risk model has not been calibrated, the explanations have not been tested with users, and adherence to legal and privacy requirements will depend on a specific assessment by the organisation.

The immediate next actions should be to build a prototype of modules 1 through 5 using synthetic telemetry; get the weights and floor values from the relevant stakeholders in security, IT, and compliance; carry out the evaluation mentioned in Section IX, including the ablation studies; perform a reviewer study to examine the usefulness of the explanations; test a minimal profile in a small organisation; compare the tier results with those of a structured assessment based on the NIST AI RMF and the applicable regulations; and, with a separate evaluation, carefully consider whether ML-based content classification should be incorporated into the design.

XII. Conclusion

The existence of Shadow AI gives rise to problems relating to enterprise governance, specifically in the areas of visibility, data protection, compliance, and accountability. Upon examining the six synopsis sources while considering the verification limitations they state, it is discovered that there is only partial, not complete and not fully integrated or validated, coverage of the elements concerning detection, risk

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

classification, explanation, and proportionate governance. The paper proposes a six-module, explainable and risk-aware framework that comprises a transparent scoring model, an uncertainty interval with an explicit indeterminate result, traceable explanations, tiered governance actions, and governed feedback, together with a threat analysis, illustrative scenarios, and a validation plan. Nevertheless, its usefulness has not yet been demonstrated: the next steps that are needed are implementation, calibration, and empirical evaluation.

References

1. S. Haag and A. Eckhardt, "Shadow IT," Bus. Inf. Syst. Eng., vol. 59, No. 6, Pp. 469-73, 2017. DOI: 10.1007/s12599-017-0497-x.
2. M. Silic and A. Back, "Shadow IT—A view from behind the curtain," Comput. Secur., vol. 45, Pp. 274-83, 2014. DOI: 10.1016/j.cose.2014.06.003
3. 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 Comput. Sci., vol. 6, Art. No. 405, 2025. DOI: 10.1007/s42979-025-03962-x
4. M. Silic, D. Silic, and K. Kind-Trüller, "From shadow IT to shadow AI—Threats, risks and opportunities for organizations," Strategic Change, 2025. DOI: 10.1002/jsc.2682.
5. A. Ojowa, M. Monteiro Marques, and A. Goncalves, "Shadow AI in organisations: A practical framework for detection, risk classification, and governance," Preprints.org, Feb. 2026. DOI: 10.20944/preprints202602.1924.v1. [Not peer reviewed].
6. S. K. Anuguthala, "Preventing shadow AI workloads at enterprise level: A novel integration of AI governance into software development lifecycle (SDLC)," Amer. Int. J. Comput. Sci. Technol. (AIJCST), WFCMLS-2026 special issue, pp. 61-75, Mar. 2026. DOI: 10.63282/3117-5481/WFCMLS26-107
7. D. Puthal, N. K. Ray, and S. P. Mohanty, "Shadow AI and the expanding attack surface: Risks, threats and countermeasures," in Proc. 2025 OITS Int. Conf. Inf. Technol. (OCIT), IEEE, 2025, pp. 1-6. [Details taken from a secondary citation; IEEE Xplore record not accessed].
8. J. A. J. Ross, L. Hibbert, and E. J. Moss, "Shadow AI: Governance, risk, and organisational resilience," in Proc. 2025 Int. Conf. Artif. Intell., Comput., Data Sci. Appl. (ACDSA), IEEE, 2025, DOI: 10.1109/ACDSA65407.2025.11166415. [Details taken from a secondary citation; full text not accessed].
9. A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion, vol. 58, Pp. 82-115, 2020. DOI: 10.1016/j.inffus.2019.12.012.
10. C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Mach. Intell., vol. 1, No. 5, Pp. 206-15, 2019. DOI: 10.1038/s42256-019-0048-x.
11. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?": Explaining the predictions of any classifier," in Proc. 22Nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135-44. DOI: 10.1145/2939672.2939778.
12. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. (neurips), vol. 30, 2017, Pp. 4765-74.
13. T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," Artif. Intell., vol. 267, Pp. 1-38, 2019. DOI: 10.1016/j.artint.2018.07.007.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

14. R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Metrics for explainable AI: Challenges and prospects," arxiv:1812.04608, 2018.
15. National Institute of Standards and Technology, "NIST Framework for Improving Critical Infrastructure Cybersecurity, Version 1.0," and "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST, Gaithersburg, MD, USA, NIST AI 100-1, Jan. 2023, DOI: 10.6028/NIST.AI.100-1.
16. National Institute of Standards and Technology, "The NIST Cybersecurity Framework (CSF) 2.0," NIST, Gaithersburg, MD, USA, NIST CSWP 29, Feb. 2024, DOI: 10.6028/NIST.CSWP.29.
17. Information Technology—Artificial intelligence — Management system, ISO/IEC Standard 42001:2023, 2023.
18. Joint Task Force Transformation Initiative, "Guide for conducting risk assessments," NIST, Gaithersburg, MD, USA, NIST SP 800-30 Rev. 1, Sep. 2012.
19. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," NIST, Gaithersburg, MD, USA, NIST SP 800-207, Aug. 2020, DOI: 10.6028/NIST.SP.800-207.
20. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Off. J. Eur. Union, L series, Jul. 12, 2024.
21. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), Off. J. Eur. Union, vol. L 119, pp. 1-88, May 4, 2016.
22. K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," J. Manage. Inf. Syst., vol. 24, No. 3, Pp. 45-77, 2007. DOI: 10.2753/MIS0742-1222240302
23. Forum of Incident Response and Security Teams (FIRST), "Common Vulnerability Scoring System v3.1: Specification document," 2019.