Position Paper: AI-Assisted Access Recertification

Rethinking Identity Governance Through Risk-Based Assurance Scoring

Author: Chiradeep Chhaya **Contact:** cbc.devel@gmail.com **Version:** 1.0 **Date:** January 2025 **Status:** For Discussion and Debate

---

Executive Summary

Access recertification—the periodic review of user entitlements to ensure appropriateness—is a foundational control in enterprise identity governance. Despite its importance, the practice has devolved into a compliance theater exercise characterized by "rubber-stamping," where reviewers approve access indiscriminately due to volume overload and lack of context.

This paper proposes a fundamental shift: from exhaustive human review of every access grant to **risk-based assurance scoring** that focuses human attention where it matters most. We present a system that calculates assurance scores based on peer typicality, resource sensitivity, and usage patterns, enabling intelligent automation of low-risk certifications while ensuring rigorous review of outliers.

Key propositions:

1. Peer typicality analysis identifies unusual access but does not determine risk; sensitivity must act as an absolute ceiling on automation.
2. Auto-certification privileges must be earned through demonstrated model performance, not administratively configured.
3. Multi-algorithm clustering with disagreement detection provides robustness against single-point-of-failure recommendations.
4. Full audit trails and three-lines-of-defense integration ensure regulatory compliance and governance oversight.

---

1. The Problem: Access Certification Has Failed

1.1 Evidence of Failure

The enterprise identity governance market has grown to address a fundamental security requirement: ensuring users have only the access they need. Access certification campaigns—periodic reviews where managers or resource owners validate entitlements—are the primary mechanism for this control.

The evidence suggests these campaigns are failing:

Metric	Finding	Source
Revocation rates	2-3% in traditional certification	Pathlock (2025)
Excessive access	Over 60% of employees have access they shouldn't	Ponemon Institute
Dormant accounts	78,000 former employees retained active credentials	Veza State of Identity Report (Dec 2025)
Permission sprawl	Average worker holds 96,000 permissions	Veza State of Identity Report (Dec 2025)

Review fatigue	Leads to "rubber-stamping" that undermines governance effectiveness	IDPro Body of Knowledge

1.2 Root Causes

Volume overload: Managers are asked to certify dozens to hundreds of access items per direct report, across multiple systems, with limited context about what each entitlement actually grants.

Lack of context: Certification interfaces typically show a list of entitlements without explaining why access might be inappropriate, what the access enables, or how the user's access compares to peers.

Misaligned incentives: Revoking access risks operational disruption and employee complaints. Approving access has no immediate negative consequence. The path of least resistance is approval.

Binary framing: Traditional certification treats all access equally—every item requires a human decision. This ignores the reality that some access is obviously appropriate (standard tools for the role) while other access warrants scrutiny (unusual privileges, dormant entitlements).

1.3 The Consequence

When certification becomes rubber-stamping, organizations lose their primary detective control for inappropriate access. Access accumulates over time (privilege creep), toxic combinations emerge undetected (segregation of duties violations), and terminated or transferred employees retain access they should no longer have.

The result: increased attack surface, compliance violations, and audit findings—precisely the outcomes certification is designed to prevent.

---

2. The Opportunity: AI-Assisted Risk-Based Certification

2.1 The Core Insight

Not all access requires equal scrutiny. A software engineer's access to their team's GitHub repository is qualitatively different from the same engineer's access to production database administration. The first is expected; the second warrants investigation.

AI and machine learning enable us to operationalize this insight at scale:

- **Peer analysis** can identify what access is typical for users in similar roles
- **Usage analytics** can identify dormant or underutilized access
- **Sensitivity classification** can identify high-risk entitlements
- **Anomaly detection** can flag outliers for focused review

2.2 Industry Movement

Major Identity Governance and Administration (IGA) vendors have introduced AI capabilities:

Vendor	Capability	Approach
SailPoint	IdentityAI, Access Insights	Peer group analysis, collaborative filtering recommendations
Saviynt	Intelligence Suite	14+ risk signals, trust scoring, peer analytics
Veza	Access AI	Graph-based analysis, natural language queries
ConductorOne	AI Copilot	Peer comparisons, risk factors, AI agents

Lumos	Agentic UARs	Autonomous AI agents for routine certifications

The market is moving toward AI-assisted certification. The question is not whether to adopt these capabilities, but how to implement them responsibly.

2.3 Regulatory Acceptance

Risk-based approaches to access review have regulatory support:

FFIEC IT Examination Handbook: "The frequency and depth of each area's audit will vary according to the risk assessment of that area."

IDPro Body of Knowledge: "Rather than reviewing all access indiscriminately, organizations should focus on high-risk permissions."

SR 11-7 / OCC 2011-12 Model Risk Management (for financial services): Joint Federal Reserve and OCC guidance providing a framework for governing AI/ML models, including validation, monitoring, and documentation requirements.

Risk-based certification is not a regulatory shortcut—it is an accepted approach when properly governed.

---

3. Our Approach: Assurance-Based Certification

3.1 Conceptual Framework

We propose an **assurance score** for each access grant that represents confidence in its appropriateness. High-assurance access can be auto-certified with audit trail; low-assurance access requires human review with AI-provided context.

```
Assurance Score = f(Peer Typicality, Resource Sensitivity, Usage Activity)
```

Peer Typicality (0-1): How common is this access among users with similar organizational placement, job function, and behavioral patterns?

Resource Sensitivity (0-1, inverted): How sensitive or critical is the resource? Higher sensitivity reduces the score.

Usage Activity (0-1): Is the access being used? Dormant access reduces the score.

3.2 Critical Design Decision: Sensitivity as a Ceiling

A naive implementation might weight these factors and allow high typicality to compensate for high sensitivity. This is dangerous.

Example of failure mode: If 80% of a department has excessive access to a critical system (perhaps due to legacy provisioning), a weighted approach would score new instances of this access as "high assurance" and auto-approve them—perpetuating the problem.

Our approach: Sensitivity acts as an **absolute ceiling**, not a weight. Access to Critical-sensitivity resources **cannot** be auto-certified regardless of typicality or usage. This ensures that human judgment remains in the loop for consequential decisions.

Sensitivity	Maximum Possible Score	Auto-Certification Eligible
Public	100	Yes

Internal	85	Yes (if graduated)
Confidential	50	No
Critical	0	Never

3.3 Peer Proximity Model

"Peer" is not a simple concept. Two users may be peers along one dimension (same team) but not another (different job function). We define peer proximity as a weighted combination of multiple dimensions:

Dimension	Weight	Factors
Structural	25%	Same manager, team, sub-LOB, LOB, location
Functional	35%	Job title similarity, job family, cost center, project assignments
Behavioral	30%	Access overlap, activity patterns, usage intensity
Temporal	10%	Tenure, time in role, hire cohort

Weights are configurable per organization to reflect local norms (e.g., matrix organizations may weight structural proximity lower).

3.4 Multi-Algorithm Clustering

Different clustering algorithms make different assumptions and produce different groupings. Rather than selecting one algorithm and accepting its blind spots, we run multiple algorithms in parallel:

- **K-means**: Centroid-based, assumes spherical clusters
- **Hierarchical**: Dendrogram-based, captures nested structures
- **DBSCAN**: Density-based, identifies natural outliers
- **Graph community detection**: Network-based, captures actual relationships

When algorithms **agree**, we have high confidence in the peer grouping. When algorithms **disagree** significantly, we flag the case for human review rather than applying a potentially incorrect recommendation.

3.5 Convergence-Based Graduation

A key innovation: auto-certification is not administratively enabled but **earned** through demonstrated performance.

Phase 1 - Observation: The model provides recommendations, but all decisions are human-made. We track:

- Recommendation acceptance rate
- Override rate
- False positive rate (flagged items that reviewers approve)

Phase 2 - Eligibility: When a category of access meets convergence criteria (e.g., >90% acceptance, <10% override, <15% false positive over 3+ campaigns), it becomes eligible for auto-certification.

Phase 3 - Graduation: A human governance review (IAM admin + compliance) approves the category for auto-certification.

Phase 4 - Monitoring: Post-graduation, metrics are monitored continuously. If thresholds are breached, the category is automatically rolled back to human review.

This approach ensures that auto-certification is based on evidence, governed by humans, and self-correcting when performance degrades.

---

4. Addressing Counterarguments

4.1 "Typical access is not necessarily safe access"

Objection: Peer analysis identifies unusual access, not risky access. A cross-trained employee or someone on a special project will appear as an outlier without being a security risk. Conversely, if everyone has excessive access, peer analysis will normalize it.

Response: We agree. This is why:

1. **Sensitivity acts as a ceiling**: High-sensitivity access cannot be auto-certified regardless of typicality.
2. **Usage is a factor**: Dormant access is flagged regardless of typicality.
3. **Peer analysis is one input, not the only input**: We combine typicality with sensitivity and usage.
4. **Legitimate outliers can be approved**: Reviewers can certify unusual access with justification, and the system learns from these decisions.

The goal is not to detect all risk—that would require threat modeling beyond certification scope. The goal is to **prioritize human attention** on items that warrant scrutiny.

4.2 "Auto-certification removes human judgment from security decisions"

Objection: Automated approval of access is inherently risky. Humans should review all access decisions.

Response: Humans already fail to review access decisions meaningfully—that's the rubber-stamping problem. A 98% approval rate with 30 seconds per decision is not "human judgment"; it's automated approval with extra steps.

Our approach:

- **Preserves human judgment where it matters**: Low-assurance and high-sensitivity access always requires human review.
- **Improves human judgment quality**: By reducing volume, reviewers can spend more time on items that warrant attention.
- **Provides better context**: AI-generated explanations help reviewers understand why access is flagged.
- **Maintains override capability**: Reviewers can always disagree with recommendations.
- **Enables second-line oversight**: Compliance can sample auto-approved items.

The alternative—requiring human review of all access—has demonstrably failed. We propose a governed alternative, not an ungoverned one.

4.3 "This shifts work from reviewers to the IAM team"

Objection: Maintaining an AI model requires significant IAM team effort. This doesn't reduce work; it moves it.

Response: Partially valid. There is work transfer:

Actor	Traditional	AI-Assisted
Reviewers	High volume, low quality decisions	Lower volume, higher quality decisions
IAM Team	Campaign administration	Model governance, validation, monitoring
Compliance	Post-hoc audit findings	Proactive sampling, threshold governance

The nature of work changes, but the net benefit depends on scale. For a 10,000-employee organization:

- Traditional: 10,000 reviewers × 50 items × 2 campaigns/year = 1,000,000 review decisions
- AI-assisted: 300,000 auto-certified + 700,000 human-reviewed (with better context)

The IAM team effort scales sublinearly while reviewer burden scales linearly. At enterprise scale, the economics favor AI assistance.

4.4 "Auditors won't accept AI-made certification decisions"

Objection: External auditors and regulators will not accept automated certification as a valid control.

Response: This depends on implementation. Auditors are unlikely to accept:

- Black-box automation without explainability
- Auto-certification without governance oversight
- Missing audit trails

Auditors are more likely to accept:

- Risk-based approaches with documented rationale (per FFIEC guidance)
- Explainable recommendations with full audit trail
- Human governance approval for automation enablement
- Second-line sampling and oversight
- Automatic rollback on control failures

Our approach is designed for the latter. We recommend engaging internal audit and compliance **before** implementation to validate acceptance.

4.5 "The model will drift as the organization changes"

Objection: Organizational changes (reorgs, M&A, new business lines) will invalidate peer groupings and cause model failure.

Response: Valid concern, which is why we include:

- **Cluster stability monitoring**: Track peer group membership changes between runs
- **Drift detection**: Alert when clustering outputs change significantly
- **Convergence-based rollback**: Categories automatically return to human review if metrics degrade
- **Revalidation on new data sources**: New systems trigger model validation (not automatic retraining)

The system is designed to fail safely—defaulting to human review when confidence is low.

---

5. Implementation Requirements

5.1 Data Requirements

Data Source	Purpose	Minimum Requirement
HR/Identity data	Organizational placement, job attributes	Current employee roster with manager, team, job code
Access data	Entitlements to certify	Access grants with resource metadata
Activity data	Usage patterns	Last-accessed timestamps, access counts

Resource metadata	Sensitivity classification	Sensitivity level per resource

5.2 Governance Requirements

Requirement	Rationale
Sensitivity classification	Resources must be classified for ceiling logic
Graduation approval process	Human sign-off required for auto-certification
Second-line sampling protocol	Compliance must sample auto-approved items
Rollback criteria	Defined thresholds for automatic rollback
Model documentation	Per SR 11-7 for financial services

5.3 Technical Requirements

Component	Options
Clustering engine	Python (scikit-learn), distributed compute for scale
Database	PostgreSQL, Snowflake for large deployments
API layer	REST API for IGA integration
Audit logging	Immutable audit trail (append-only)

---

6. Success Metrics

6.1 Efficiency Metrics

Metric	Baseline	Target	Measurement
Time per decision	Varies	50% reduction	Time tracking in UI
Campaign completion time	Weeks	Days	Campaign duration
Review volume per reviewer	100% of items	30-40%	Items requiring human review

6.2 Quality Metrics

Metric	Baseline	Target	Measurement
Revocation rate	2-3%	10-15%	Revocations / total decisions
False positive rate	N/A	<20%	Flagged items approved by reviewer
Override rate	N/A	<15%	Recommendations overridden

6.3 Compliance Metrics

Metric	Target	Measurement

Certification completion	>95%	Completed / assigned items
Audit findings (identity-related)	80% reduction	Year-over-year comparison
Auto-certification sampling	100% coverage	Sampled / auto-certified

---

7. Risks and Mitigations

Risk	Likelihood	Impact	Mitigation
False negatives (risky access auto-approved)	Medium	High	Sensitivity ceiling, dormancy detection, second-line sampling
False positives (legitimate access flagged)	Medium	Low	Reviewers can approve with justification; model learns
Model drift	Medium	Medium	Continuous monitoring, automatic rollback
Regulatory rejection	Low	High	Pre-implementation engagement with audit/compliance
Over-reliance on AI	Medium	Medium	Mandatory human review for high-sensitivity; caps on auto-cert rate
Implementation complexity	Medium	Medium	Phased rollout, starting with recommendation-only

---

8. Conclusion

Access recertification in its current form has failed. The evidence is clear: near-universal approval rates indicate rubber-stamping, not meaningful review. This failure creates real security risk—privilege creep, toxic combinations, and dormant access that attackers can exploit.

AI-assisted certification offers a path forward, but implementation matters. Naive approaches—weighted risk scores, administratively configured auto-approval, single-algorithm recommendations—create new risks while solving old ones.

We propose a **governed, evidence-based approach**:

1. **Assurance scoring** that combines peer typicality, resource sensitivity, and usage activity
2. **Sensitivity as a ceiling** that prevents auto-certification of high-risk access regardless of typicality
3. **Convergence-based graduation** that requires demonstrated performance before enabling automation
4. **Multi-algorithm robustness** that flags disagreement rather than forcing a single answer
5. **Three lines of defense** integration that maintains governance oversight

This approach does not eliminate human judgment—it focuses human judgment where it matters and provides the context to make that judgment meaningful.

The question is not whether to adopt AI-assisted certification. The market is moving in this direction. The question is whether to implement it thoughtfully, with appropriate governance and safeguards, or to deploy it hastily and create new categories of risk.

We advocate for the former.

Appendix A: Literature References

Academic Research

- Bolton, R.J. & Hand, D.J. (2001). "Peer Group Analysis - Local Anomaly Detection in Longitudinal Data." https://www.researchgate.net/publication/2484688_Peer_Group_Analysis_--_Local_Anomaly_Detection_in_Longitudinal_Data

- Molloy, I., Li, N., Li, T., Mao, Z., Wang, Q. & Lobo, J. (2009). "Evaluating Role Mining Algorithms." SACMAT'09, Purdue University / IBM T.J. Watson Research Center. https://www.cs.purdue.edu/homes/ninghui/papers/eval_role_mining_sacmat09.pdf

Regulatory Guidance

- Federal Financial Institutions Examination Council. "IT Examination Handbook: Risk Assessment and Risk-Based Auditing." https://ithandbook.ffiec.gov/it-booklets/audit/risk-assessment-and-risk-based-auditing

- Federal Reserve Board & Office of the Comptroller of the Currency. "SR 11-7 / OCC 2011-12: Supervisory Guidance on Model Risk Management." (April 2011) https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm

- U.S. Congress. Sarbanes-Oxley Act, Section 404: Management Assessment of Internal Controls.

- European Union. General Data Protection Regulation, Articles 5 and 22.

Industry Research

- IDPro Body of Knowledge. "Optimizing Access Recertifications." V. Gupta (2025). https://bok.idpro.org/article/id/119/

- Veza. "State of Identity & Access Research Report." (December 2025) https://veza.com/company/press-room/veza-identity-access-research-report-reveals-identity-permissions-sprawl-has-reached-critical-levels-amid-explosion-of-machine-and-ai-agent-identities-across-the-enterprise/

- Pathlock. "2025 Digital Transformation and Access Risk Report." https://pathlock.com/simplify-identity-governance-in-a-multi-application-world/

- Ponemon Institute. "Privileged Access Management Study." https://www.ponemon.org/news-updates/blog/security/new-research-on-privileged-access-management-reveals-the-status-quo-is-not-secure.html

- U.S. Treasury / CFIUS. "T-Mobile $60 Million Penalty." (August 2024) https://home.treasury.gov/policy-issues/international/the-committee-on-foreign-investment-in-the-united-states-cfius/cfius-enforcement

- NYDFS. "23 NYCRR Part 500 Cybersecurity Regulation - Penalty Structure." https://www.dfs.ny.gov/industry_guidance/cybersecurity

Vendor Documentation

- SailPoint. "IdentityAI and Access Insights Documentation." AI capabilities since 2017. https://www.sailpoint.com/products/identity-ai

- Saviynt. "Intelligence Suite Documentation." 14+ risk signals, trust scoring. https://saviynt.com/intelligence

- Veza. "Access AI Documentation." https://veza.com/products/access-ai/

- ConductorOne. "AI Copilot Documentation." https://www.conductorone.com/products/ai-native-identity/

Appendix B: Glossary

Term	Definition
Access certification	Periodic review of user entitlements to validate appropriateness
Assurance score	Numeric representation of confidence in access appropriateness
Auto-certification	Automated approval of access without human review
Convergence	State where model recommendations consistently match human decisions
Dormant access	Entitlements that exist but are not used
Graduation	Process by which an access category earns auto-certification privileges
IGA	Identity Governance and Administration
Peer typicality	Degree to which access is common among similar users
Privilege creep	Accumulation of access over time beyond job requirements
Rubber-stamping	Approving access without meaningful review
Sensitivity ceiling	Design pattern where resource sensitivity limits automation eligibility
SoD	Segregation of Duties (conflicting access that should not coexist)

Appendix C: Revision History

Version	Date	Author	Changes
1.0	January 2025	Chiradeep Chhaya	Initial draft for stakeholder review

---

This document is intended for discussion and debate. Comments and counterarguments are welcome.