PROTECTING THE ORGANIZATION FROM WITHIN: AUGMENTING HUMAN JUDGMENT WITH BEHAVIORAL ANALYTICS

AI for Insider Threat Detection

Executive briefing — Security operations and human resources risk leaders | ~20 min

Why Insider Threat Deserves Board-Level Attention

- Insider incidents are harder to detect than external attacks because the actor already has legitimate access
- Industry-reported ranges suggest insider incidents take substantially longer to identify and contain than external breaches — treat as directional, not a specific verified figure
- Impact spans data loss, regulatory exposure, and reputational harm, not just direct financial cost
- AI-assisted detection does not replace HR and security judgment — it surfaces signal earlier for humans to act on
- This briefing outlines the technical approach, the guardrails required, and a phased path to adoption

Three Categories of Insider Risk

- Malicious insiders: employees or contractors who deliberately misuse access for personal gain, sabotage, or on behalf of a third party
- Negligent insiders: well-intentioned staff who create risk through careless handling of data, weak security hygiene, or policy shortcuts
- Compromised credentials: an outsider operating through a legitimate employee's stolen or phished account, indistinguishable from that user at the access-control layer
- Each category requires a different detection signature and a different response playbook
- Most programs underinvest in the negligent category despite it representing a large share of day-to-day incidents in industry reporting

User and Entity Behavior Analytics (UEBA): The Core Technique

- UEBA builds a statistical profile of normal behavior for each user, device, and service account, then scores deviations from that baseline
- It correlates signals across systems — file access, login patterns, email, endpoint, cloud apps — rather than relying on any single log source
- Machine learning models flag anomalies in context (peer group, role, time, location) rather than against fixed rule thresholds
- This reduces reliance on static rules that miss novel behavior and generate excessive noise
- UEBA output is a risk score and evidence trail for analysts — not an automated accusation or an automated action

Establishing the Behavioral Baseline

- Baselines are built per individual and per peer group (role, team, department) to account for legitimate variation in normal work
- A minimum observation window (typically several weeks) is needed before a baseline is considered reliable
- Baselines must be refreshed on a schedule to reflect role changes, seasonal work patterns, and organizational shifts
- Static baselines set once and left unmaintained become a source of false confidence, not protection
- Baseline quality depends directly on data source coverage and consistency — gaps in logging create blind spots

Detecting Data Exfiltration Patterns

- Common patterns include unusual volume or timing of downloads, bulk access to files outside a user's normal scope, and transfers to unsanctioned personal cloud storage
- Sequence matters more than any single event: e.g., mass file access followed by compression followed by an external transfer, viewed together
- Off-hours activity and access immediately preceding a resignation or role change are elevated-risk indicators, not automatic proof of intent
- Illustrative scenario, not a verified case study: an employee's outbound data volume to a personal email account rises sharply in the two weeks before their resignation date — a pattern designed to prompt review, not automatic escalation
- Detection should weight patterns and context together — a single anomalous event is rarely sufficient grounds for action

Privileged User Monitoring

- Privileged accounts (admins, database owners, executives with broad access) carry outsized risk and warrant a higher standard of monitoring
- Session recording, just-in-time access, and step-up authentication reduce the window of unmonitored privileged activity
- Separation of duties should apply to the monitoring function itself — the team reviewing privileged activity should not be the same team holding those privileges unchecked
- Privileged monitoring is the area most likely to require explicit union or works-council consultation given its intrusiveness
- Treat privileged monitoring as a distinct workstream with its own approval chain, not an extension of general employee monitoring

Balancing Detection With Employee Privacy and Trust

- Overly intrusive monitoring can suppress the psychological safety that makes employees willing to report concerns or admit mistakes
- Purpose limitation matters: collect and use behavioral data specifically for security risk, not general performance surveillance
- Transparency about what is monitored (without disclosing detection thresholds) tends to support both trust and deterrence
- Data minimization — collecting only what is proportionate to the risk being addressed — should be a design principle, not an afterthought
- Employee communication and policy acknowledgment are as important to program success as the technology itself

False Positive Management and Avoiding Over-Surveillance Culture

- High false-positive rates erode analyst trust in the system and can lead to alert fatigue or premature dismissal of real signals
- Tuning models to organizational context (roles, seasonal cycles, business travel) reduces noise without reducing coverage
- A tiered alerting model — low-confidence signals routed to automated review, high-confidence signals routed to analysts — helps manage volume
- Every escalation to HR or an employee-facing action should require a documented, reviewable rationale, not a raw score alone
- Track false-positive and false-negative rates over time as a formal program metric, not just detection volume

Integrating With HR Processes as Detection Triggers

- Offboarding is a known high-risk window: access should be provisioned to revoke automatically and behavioral monitoring should intensify in the departure period
- Role changes and internal transfers should trigger a re-baseline and an access review, since entitlement creep is a common source of unnecessary risk
- Performance management events (e.g., disciplinary action, layoff notice) are contextual risk factors that HR is positioned to flag to security under agreed protocols
- This requires a formal, documented handoff process between HR and security — not ad hoc or informal information sharing
- Any HR-security data sharing arrangement should be scoped, logged, and reviewed periodically for continued necessity

Legal and Works-Council Considerations

- Employee monitoring is subject to jurisdiction-specific data protection, labor, and privacy law that varies significantly by country and region
- In many jurisdictions, works councils or employee representative bodies have consultation or co-determination rights over monitoring programs — engage early, not after deployment
- This briefing provides general practice guidance only and is not a substitute for legal advice from qualified counsel in each relevant jurisdiction
- Cross-border organizations should expect to run different monitoring configurations by location rather than a single global policy
- Legal review should be a gating step before any new data source or detection capability goes live, not a retrospective check

Governance: Who Can Access Insider Threat Data

- Access to raw behavioral data and case files should be restricted to a named, limited group under documented need-to-know
- A cross-functional review board (security, HR, legal, and where applicable employee representatives) should approve escalations before employee-facing action is taken
- Audit logging of who accessed insider threat data, when, and why should itself be monitored — this is a high-sensitivity data set
- Retention limits should be defined and enforced so behavioral data is not held indefinitely beyond its operational purpose
- Regular independent review of the program's access controls and case outcomes helps sustain internal and external credibility

Program Governance Model

- A charter should define the program's purpose, scope, and explicit boundaries (what it will and will not be used for)
- Ownership sits jointly across security, HR, and legal — no single function should govern the program alone
- Escalation pathways should be pre-agreed and documented so response is consistent rather than improvised case by case
- Periodic external or internal audit of the program builds the accountability needed to sustain employee and works-council trust
- Governance should be revisited on a fixed cadence as regulations, tooling, and organizational structure evolve

Phased Program Rollout

- Phase 1 (0–3 months): stand up data source integration, establish legal/works-council engagement, and define governance charter
- Phase 2 (3–6 months): deploy UEBA in monitoring-only mode to build and validate baselines without triggering employee-facing action
- Phase 3 (6–9 months): introduce tiered alerting and analyst workflows with a small, well-defined use case such as offboarding-window monitoring
- Phase 4 (9–12 months): expand coverage to privileged users and additional data sources, with false-positive rates and case outcomes reviewed at each gate
- Each phase gate requires sign-off from security, HR, and legal before expanding scope or data sources

Next Steps and the Ask

- Approve formation of the cross-functional governance board (security, HR, legal) to own charter development
- Commit to Phase 1 scope: data source inventory, legal and works-council engagement, and baseline governance charter, targeted for the next quarter
- Identify a bounded pilot use case (recommended: offboarding-window monitoring) to prove value before broader rollout
- Allocate budget and staffing for monitoring-only Phase 2 before any employee-facing escalation capability is enabled
- Schedule a follow-up review in 90 days to assess Phase 1 completion and confirm readiness to proceed to Phase 2