

Transparent Risk Scoring Across Netskope One

The design philosophy, evolution, mechanics, and interoperability of Netskope's Cloud Confidence Index, User Confidence Index, MCP Confidence Level, and AI SecOps agent scores; and how any buyer can verify every claim.



Table of Contents

EXECUTIVE OVERVIEW: RISK SCORES ARE THE DECISION LAYER	3
SCORING AS PLATFORM INFRASTRUCTURE	3
SIX PRINCIPLES BEHIND EVERY NETSKOPE SCORE	4
HOW THE PORTFOLIO GREW	5
THE FOUR SCORING SYSTEMS TODAY	7
CLOUD CONFIDENCE INDEX: SCORING APPLICATIONS	8
USER CONFIDENCE INDEX: SCORING IDENTITIES	9
MCP CONFIDENCE LEVEL: SCORING THE AGENTIC INTEGRATION LAYER	10
AISECOPS AGENT SCORES: DETERMINISTIC FLOORS, BOUNDED MODELS	12
INTEROPERABILITY: A FABRIC OF SCORES, NOT A HIERARCHY	13
SCORES AS TIME SERIES: THE TEMPORAL ARCHITECTURE	15
THE FIVE BUYER QUESTIONS	17
THE NEXT FIVE QUESTIONS (AND NETSKOPE'S CURRENT ANSWERS)	18
LIMITATIONS AND HONEST BOUNDARIES	19
SCORES ARE JUDGMENT MADE LEGIBLE	20
REFERENCES	20

EXECUTIVE OVERVIEW: RISK SCORES ARE THE DECISION LAYER

For modern security platforms, risk scores are the decision layer. They score which applications should be approved, which users deserve investigation, which Model Context Protocol (MCP) servers an agent may reach, and which incidents an analyst opens first.

Measurement of these platforms is now included in many RFPs to help buyer teams evaluate the hard numbers behind the commercial claims. These are worthwhile and valuable, and Netskope would like to see them go further.

This paper is a complete account of that decision layer within Netskope One. It begins with the design philosophy: Six principles that every Netskope scorer obeys, from one-entity-one-scale to deterministic-first computation, and traces the portfolio expansion across four eras of enterprise risk, from SaaS sprawl through insider threat to agentic AI and finally to scoring the operational queue itself.

The paper then documents each system at the level of inputs, formulas, scales, and controls: the Cloud Confidence Index (CCI) for applications, the User Confidence Index (UCI) for identities, the MCP Confidence Level for MCP servers, and the AISEcOps agent scores for cases, watchlist users, and AI entities.

Two further sections examine properties that single-system descriptions miss, namely the interoperability fabric through which scorers exchange signals along deliberately one-directional paths, and the temporal architecture that makes trends, decay curves, and evidence windows scored quantities in their own right.

The paper closes with the transparency frame that now appears regularly in RFP criteria: Five questions every buyer should ask of any vendor-supplied score. It answers these questions for all four systems, and adds five further questions that Netskope proposes the industry adopt.

Scope and Method

This paper draws exclusively on public Netskope Knowledge Portal documentation for CCI, UCI, and MCP scoring, and on the technical brief Risk Scoring in Netskope AISEcOps (July 2026) for the agent scores. No mechanism is described here that a customer cannot verify in a licensed tenant. Features noted as requiring additional licensing are identified as such.

SCORING AS PLATFORM INFRASTRUCTURE

Every ranking decision is a score, whether or not the vendor calls it one

Every consequential action a security platform takes (to allow, coach, quarantine, investigate, or escalate) is preceded by a ranking decision, and every ranking decision is a score regardless of whether the vendor calls it one.

Treated casually, scores accumulate as point-product afterthoughts: incompatible scales, undocumented weights, thresholds set once, and never revisited. Treated as infrastructure, they become a coherent decision layer with a shared design grammar, published mechanics, and defined interfaces between systems. Netskope has built risk scores as a coherent decision layer through more than a decade of platform evolution.

Three developments have made scoring architecture both a board and procurement concern.

Firstly, scores now drive automated enforcement: A number that gates real-time policy must withstand engineering scrutiny, not just marketing review.

Secondly, AI has multiplied the entities being scored. That includes applications, embedded AI features, autonomous agents, MCP servers, and the humans operating them, so a single miscalibrated scale propagates further than it once did.

Thirdly, the scoring vocabulary itself has degraded across the industry, with confidence, severity, and likelihood frequently blended into single numbers and unrecoverable statistical meaning.

In response, industry analysts have equipped buyers with pointed questions about any vendor-supplied score. Buyers must ask how it is calculated, what grounded truth it is calibrated against, how fresh it is, whether it can be reweighted, and who set the thresholds. Netskope's position is that this scrutiny is warranted, and that the correct response is transparent publication, not merely reassurance.

SIX PRINCIPLES BEHIND EVERY NETSKOPE SCORE

Scoring systems have at least four failure routes:



The blended super-score nobody can decompose



The vendor default nobody can adjust



The model output nobody can bound



The number that moved and nobody can say why

Netskope's portfolio was built to avoid them all.

Six principles make the portfolio an architecture rather than an accumulation.

1. **One entity, one question, one scale**

Application readiness, user trust, MCP integration risk, and investigative priority are four different questions, so they get four scales, not one blended number. (The Insider Risk agent even returns a level and a verdict, because "how urgent" and "what was concluded" differ.)

2. **Deterministic first; models argue inside bounds**

Arithmetic sets the score or its floor before any model runs. Where a model participates, it works within published limits. No amount of adversarial text in a filename or prompt can push a score below a deterministic floor.

3. **Evidence travels with every number**

Each score arrives with its decomposition (factor contributions, signal trails, sub-scores, per-attribute confidence). Deterministic components emit their own execution record, so the explanation cannot drift from the computation.

4. Research sets the bar

The tenant holds the dials. Netskope research owns the defaults, because they demand evidence at a scale no single tenant sees. Every system then exposes a tenant-side control (CCI sliders, editable UCI impacts, MCP overrides, and natural-language guidance for the AISEcOps agents) with deterministic floors fixed beneath all of them.

5. Scores move, and the record of movement stays

A current score without its history is an assertion about the past dressed as the present. Every movement, be it deduction, third-party ingestion, revert, reset, or re-investigation, is evented, timestamped, and inspectable.

6. “Unknown” is never treated as safe

Unrated entities get explicit conservative defaults (an uncataloged MCP server scores 80/100; a missing UCI anchors at medium; absent telemetry yields inconclusive, never legitimate), and the system refuses to invent signals it does not have.

HOW THE PORTFOLIO GREW

Like almost every other portfolio, Netskope’s was not designed in one sitting, and its history is worth telling because the pattern of expansion is itself an argument for the architecture. Each scorer arrived when a new class of entity began making or mediating decisions about enterprise data. And each arrival reused the same design grammar. Figure 1 summarizes the arc.

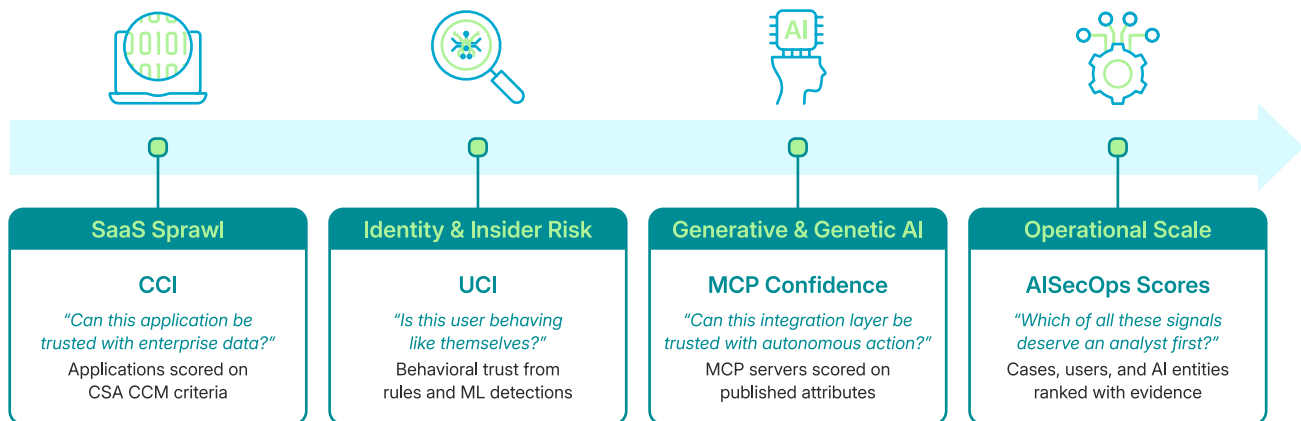


Figure 1: Four eras of scoring expansion.

The first page: Application sprawl

When SaaS adoption outran procurement, the governing question was whether a given application could be trusted with enterprise data at all, and CCI answered it with objective criteria adapted from the CSA Cloud Controls Matrix over a catalog that now exceeds 64,000 applications.

The second era: Introspection

Credential compromise and insider risk made the uncertain party the user rather than the application, and UCI answered with a behavioral trust score fed first by sequential rules and later by machine-learned detectors as the user and entity behavior analytics (UEBA) discipline matured. The expansion from basic to advanced user behavior analytics (UBA) was visible in the product's own licensing tiers.

The third era: AI moves the goal posts

AI has moved the perimeter from applications and users to the integration layer between them: MCP servers and the agents that call them can read, write, and delete on a user's behalf. As a result, the governing question has become whether an integration layer can be trusted with autonomous action. The MCP Confidence Level answered this question with attribute-based composites over published evidence, inventoried alongside the AI applications and identities discovered by the platform's AI inventory.

Era four: AI SecOps steps in

The fourth expansion is the most recent, and the most self-referential, and is concerned with tackling signal overload. The preceding three eras started to produce more scored signals than human teams can triage, so AI SecOps scores the operational queue itself, sorting which cases, which watchlist users, which AI entities deserve an analyst first, with agents whose own scoring obeys the same grammar it ranks.

The portfolio's growth, in other words, tracks the growth of the enterprise attack surface, and the constant across all four eras is not any formula but the principles the formulas obey.

"The constant across all four eras of the Netskope portfolio is not any formula, but the principles the formulas obey."

THE FOUR SCORING SYSTEMS TODAY

Scheduled cadence, separated from event-driven movement triggers

Table 1 summarizes the four systems as they stand today, separating scheduled cadence from movement triggers. A score refreshed on a schedule answers the question of how fresh the data needs to be, while a score moved by events addresses the need to understand what activity drives change.

SYSTEM	SCORED ENTITY	SCALE AND BANDS	CADENCE	MOVEMENT TRIGGERS	CUSTOMER CONTROLS
Cloud Confidence Index (CCI)	Cloud and AI applications (64,000+ cataloged)	0-100; Five Cloud Confidence Levels; Poor, low, medium, high, excellent	Continuous search program	New evidence from vendors or observation; customer assessment requirements.	Per-app risk-weight sliders across seven categories; seven categories; reset to defaults
User Confidence Index (UCI)	Enterprise users	1,000+ (max trust) descending; lower is riskier	Event-driven; no batch schedule	Each rule or ML detection (published impact brands); third-party API impact; revert; audited reset	Per-policy score impacts editable; REST APIs for impact, revert, reset
MCP Confidence Level	Public MCP servers	0-100 composite; banded confidence levels (e.g., Excellent 90-100)	Maintained in the catalog	Provider-published attribute changes; customer recalculation (point-in-time)	Per-attribute override, on-demand recalculation, reset
AI SecOps agent scores	DLP cases; watchlist users; users, AI apps, and MCP servers	Case: 0.0-1.0, four bands. Insider: four levels+ verdict. AI Risk: 0-100 per entity type	Insider and AI Risk: daily. Cases: on creation	Substantive new case evidence (re-investigation, capped daily); watchlist changes; UCI movement	Natural-language guidance retained in agent memory; deterministic floors fixed; analyst confirm, dismiss, adjust

Table 1: The four scoring systems of the Netskope One platform.

CLOUD CONFIDENCE INDEX: SCORING APPLICATIONS

Objective criteria adapted from the CSA Cloud Controls Matrix

Inputs and computation

TCCI is a database of cloud and AI applications evaluated against objective criteria adapted from the Cloud Security Alliance (CSA) Cloud Controls Matrix (CCM), the CSA's cybersecurity control framework for cloud computing, measuring enterprise readiness across security, auditability, and business continuity.

Two input classes feed each evaluation: Structured responses from application vendors and capabilities observed directly by Netskope research. All inputs are transformed to numeric form and combined by a learning algorithm into a score normalized between 0 and 100. This places the application into one of five Cloud Confidence Levels (CCL): Poor, Low, Medium, High, or Excellent.

On top of the normalized combination, Netskope applies rewards and penalties at the category level. Each attribute an application satisfies earns a reward; each attribute it fails incurs a penalty. The magnitude of both is set per application category rather than per application, which keeps the treatment consistent and auditable across the catalog.

The category context matters because for a cloud storage application, failing to encrypt data at rest carries a severe penalty and satisfying it earns little reward, because encryption-at-rest is table stakes for that category. For a social application the same attribute is weighted less severely. The category-level reward and penalty vectors are combined with the application's attribute vector and the result is normalized to the 0-100 scale, so an application's score is reproducible from its attribute values and its category's published weighting posture.

Reweighting to the buyer's environment

Administrators can customize CCI attribute weights per application through Risk Weight Settings, adjusting slider weights across seven categories:



Each contains sub-categories specific to each application. Adjusted weights recompute the effective assessment for that tenant, and a Reset Risk Weights control returns the defaults, so customization is reversible and the vendor baseline is never lost. This capability requires additional licensing.

Applications absent from the catalog can be submitted for evaluation by the CCI research team directly from the console, which is also the mechanism by which refresh requests enter the research queue.

USER CONFIDENCE INDEX: SCORING IDENTITIES

Behavioral trust that complements content inspection

Where the Netskope One Data Loss Prevention (DLP) engine inspects content (what data is moving, and whether policy permits it) UCI inspects behavior, identifying evasive patterns of a malicious insider that no single content event reveals. That behavior might be the slow escalation, the unusual destination, or the trust that erodes across a sequence of otherwise-permitted actions. The two are complementary controls.

Combined in a real-time policy, DLP's content inspection and UCI's behavioral trust score form a comprehensive data protection posture. Together, they can block a transfer based on content and independently constrain a user whose behavior has turned evasive, even when individual actions clear content checks.

Detections, severities, and score impacts

UCI is a per-user trust score beginning at 1,000, where a lower number means a riskier user. This score is reduced by two detection families: predefined and custom sequential rules, and machine-learned models covering suspicious insider behavior, suspicious network access, and compromised-device indicators (the ML family requires Advanced UBA licensing).

Every detection carries a severity with a fixed score impact band: Critical detections impact the score by 251 to 350 points, High by 151 to 250, Medium by 101 to 150, Low by 51 to 100, and Informational by 1 to 50. The bands answer the thresholds question directly. An administrator can read, per policy, exactly how much trust a given detection removes. Policies can be edited, cloned, or disabled, with a pending-changes review showing every modification and its author before it is applied.

An open scoring surface: programmatic impact revert, and reset

UCI is not a closed loop. The UCI Impact REST API allows third-party systems, such as an endpoint detection platform, identity provider, or an email security gateway, to reduce a user's UCI by a specified amount. The user, source, timestamp, score reduction, and reason are recorded as parameters. Each ingestion generates an anomaly with a durable identifier, retrievable individually or in bulk for 90-days. Because third-party judgments can be wrong, the ingestion can be completely reversed, 'Mark Anomaly as Allowed API' restores the deducted trust while preserving the alert record for audit. Finally, the UCI Reset API returns users to the 1,000-point maximum, and the reset itself is written to the incident record as an admin-attributed anomaly. Every movement of the number (down, back up, or to baseline) leaves a trail that can be inspected in behavior analytics.

"A score that can only fall converges on pessimism. The paired Impact and Allow APIs mean UCI can absorb external risk signals without permanently inheriting their false positives."

MCP CONFIDENCE LEVEL: SCORING THE AGENTIC INTEGRATION LAYER

Attribute-based composites over published evidence

MCP servers are the integration and execution layer for agentic AI, and they concentrate a novel risk shape. A widely adopted server with write-capable tools has a blast radius unlike a conventional SaaS application. The MCP Catalog, reached under the Generative AI section of the App Catalog, inventories and scores publicly-available MCP servers (both official servers hosted by vendors and community-developed code repositories). The resulting score is a weighted average computed over risk attributes grouped into families:

- Authentication (supported grant types from OAuth authorization-code flow down to static API tokens)
- Data Protection (transport encryption)
- Deployment Model (local, container, remote)
- Destructive Tools (whether any tool performs update or delete operations)
- Main Characteristics (official versus community classification, protocol version, pricing, hosting provider)
- Protocols Support

Each attribute carries a value and an associated confidence level, and the catalog states its evidentiary basis plainly. Attributes are based on information published by the MCP server provider. The composite maps to confidence bands whose boundaries are displayed in the interface itself: Excellent, for example, is 90 to 100.

Non-read-only capability is surfaced as a first class attribute because it categorically changes the authorization decision. An organization may reasonably approve read-only servers broadly, while requiring individual review for anything that can update or delete. Catalog data, including the scores, feeds real-time access control policy directly.

FAMILY	ATTRIBUTE	VALUE	CONFIDENCE
Authentication	OAuth 2.0 authorization-code grant	Yes	High
	OAuth 2.0 client-credentials grant / bearer / PAT / API token	Not reported	—
Data Protection	TLS support	Not reported	—
Deployment Model	Remote	Yes	High
	Local / container	Not reported	—
Destructive Tools	Update operations / delete operations	Not reported	—
Main Characteristics	Classification	Official	High
	MCP protocol version	2025-06-18	High
	Pricing	Free	Low
	Hosting provider	SaaS vendor	High
Protocols Support	HTTP / SSE	Yes	Medium
	STDIO / streamable HTTP / FastAPI	Not reported	—

Table 3: A representative MCP server risk-attribute breakdown. "Not reported" attributes carry no confidence and do not contribute; the composite is the weighted average of the attributes the provider has published.

Recalculation in the buyer's hands

The catalog answers the reweighting question with a product control rather than a services engagement. From any server's detail page, Configure Attributes lets an administrator override attribute values to reflect their own deployment reality; disabling, for example, an authentication method their tenant will not use. The Calculate action then recomputes the confidence score from the modified attribute set, displayed alongside the unmodified default.

The documentation is precise about epistemic status. The recalculated score is a point-in-time computation over the values the administrator chose, an aid to an authorization decision rather than a replacement for the maintained default. A Reset restores the original attributes and score. As a result, the default and the customization can always be told apart, which is the property that keeps tenant-specific tuning from silently drifting away from the vendor baseline.

AISECOPS AGENT SCORES: DETERMINISTIC FLOORS, BOUNDED MODELS

One discipline, three agents

The AISECOPS layer applies three AI agents to security operations, and the scoring discipline is the same in all three. Deterministic computation establishes the score or its floor before any model is involved. Where a large language model participates, it works inside published bounds. The DLP agent scores cases assembled from clustered DLP incidents on a 0.0 to 1.0 scale using a fixed, weighted, additive rubric of six factors. Data sensitivity carries the largest weight at 0.30, followed by incident severity at 0.25, trust trend and incident velocity at 0.15 each, and trust level and match volume at 0.10 each. Every contribution tier is published, hard floors that no factor sum can undercut are established (confirmed credentials moving externally floor a case at high), along with a structured mitigation sub-rubric for data that belongs to the user personally. Band thresholds sit at 0.75, 0.50, and 0.25 and live in configuration, not in the model, so tuning them requires no retraining and no prompt change.

The Insider Risk agent evaluates watchlist users daily over a 30 day evidence window with 90 days of trust history. A deterministic calculator anchors a four level baseline to the user's UCI band and applies published escalation rules and floors; the model may then escalate the level freely but may lower it only with cited justification, and is barred from lowering it at all while critical DLP findings remain unexplained or an offboarding-monitored user shows movement toward personal or AI destinations.

The AI Risk agent, which ranks users, generative AI applications, and MCP servers by AI specific exposure, is computed entirely with deterministic arithmetic over published per category formulas and weights. The language model never touches the number and only writes the narrative for the top ranked entities. All three deliver factor-level breakdowns with every score.

Steering the agents: natural language, not a weights console

The other three scoring systems expose their tuning as configuration controls (CCI sliders, editable UCI impacts, MCP attribute overrides) because they are deterministic engines with fixed knobs. The AISECOPS scorers are different because they are agents, so the natural way to influence them is the way one directs an analyst; in words.

Rather than expose a weights console, AISECOPS lets a security team record standing guidance in natural language. For example, the security team can note that a particular sanctioned destination is expected for a given team, that a class of activity is routine in their environment, or that a certain data pattern warrants heightened concern. That guidance is retained in the agent's memory and applied on subsequent investigations, shaping how evidence is weighed and how cases are dispositioned without a model retrain or a configuration change. The deterministic floors and published band cutoffs still hold underneath. Guidance steers judgment within the bounds, it does not lower a floor, so natural-language influence and deterministic safety co-exist.

Calibration and ground truth

Grounded truth in security scoring is adjudication: What did a qualified analyst conclude when they actually investigated? The AISECops design treats analyst dispositions as labeled outcomes. Confirmations and dismissals are persisted alongside the factor sub-scores that produced each score, giving Netskope a basis for re-tuning deterministic weights offline against real investigative results. Tenants with sufficient adjudicated history run the DLP triage gate on a model trained on their own outcomes rather than the global baseline.

Band thresholds for the AI Risk agent were set against production snapshots with a stated calibration target (a handful of critical applications per tenant, not dozens) and the internal evaluation history is candid about failure modes. In this case, an unconstrained model was observed rating nearly every watchlist user low risk, and the symmetric burden of proof on escalation and de-escalation was added specifically to correct that bias.

Limits are stated simply. The agents see only what platform enforcement points see, unselected cases remain unscored rather than defaulted, and a failed model call surfaces as a failure, never an invented number.

INTEROPERABILITY: A FABRIC OF SCORES, NOT A HIERARCHY

One-directional signal paths, with output-to-input loops explicitly cut

One-directional paths, explicitly cut loops

Four scoring systems that never communicated would waste each other's evidence. Four that freely consumed each other's outputs would amplify each other's errors. The Netskope portfolio takes a third position: Scores flow along defined, one-directional paths, and wherever an output could re-enter as an input, the loop is cut explicitly.

Reading the graphic

UCI informs all three AISECops agents: as a rubric factor and trend in the DLP agent, the baseline anchor in the Insider Risk agent, and an inverted component of user context in the AI Risk composite. CCI supplies application posture to the AI Risk agent (computed as 100 minus the CCI, with CCL-based penalties and credits) and, with UCI and MCP Confidence, grounds real-time policy. The MCP Catalog supplies server posture. DLP case history feeds insider watchlist selection and the AI Risk user category, and insider watchlist standing feeds the AI Risk user score. But watchlist selection counts only DLP cases, never cases the Insider or AI Risk agents themselves created, which keeps the selection from feeding on its own output. Three scores corroborate one another without amplifying one another.

Open at the edges: signals in, decisions out

Interoperability extends beyond Netskope's own systems in both directions. Inbound, the UCI Impact API (reachable directly or through Cloud Risk Exchange integrations) lets endpoint, identity, and email security platforms such as CrowdStrike, Microsoft Defender, SentinelOne, and Mimecast contribute risk impact to a user's UCI, with every contribution recorded as a searchable anomaly and every contribution revertible through the Allow API. This is how customers extend realtime zero trust policy: on the strength of signals suggesting a malicious insider detected in those other systems. An endpoint or email platform flags evasive behavior, the signal lowers the user's UCI, and a UCI-based policy tightens access in the same enforcement path that governs approved applications, without the security team stitching the systems together by hand.

A partner's judgment can move the score, but it cannot move it silently, and it cannot move it permanently without surviving review. Outbound, scores become decisions. CCL levels, UCI bands, and MCP confidence data drive real-time policy, so a score is not a report an administrator reads but a control surface the platform enforces. And where the same entity is visible in two places consistency is structural rather than aspirational: AI Risk results publish back to the AI Command Center, so the risk label an analyst sees in the inventory matches the one in AISecOps by construction, not by synchronization effort.

The fabric is also why the portfolio resists the single-blended-score temptation. Because each scorer owns one question and exports its answer along known paths, a consuming system can weight that answer for its own purpose without the source system needing to anticipate every consumer. For example, the AI Risk agent inverts UCI at half weight inside one sub-category. A universal score would have to bake those weightings in once, for everyone, forever.

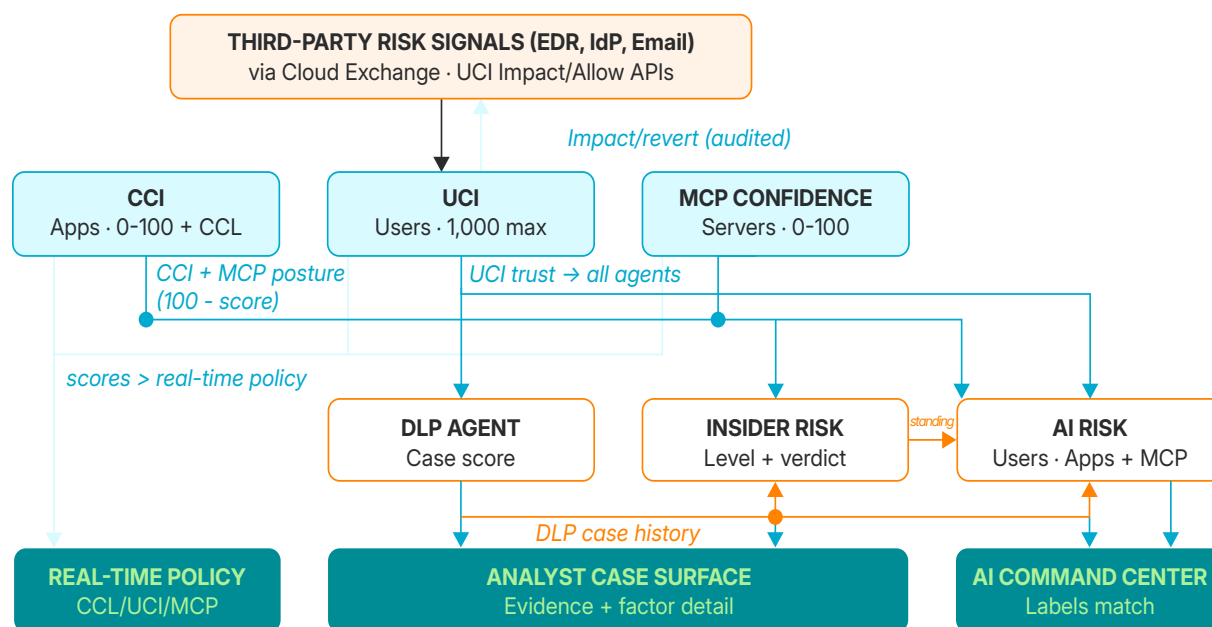


Figure 2: The "The interoperability fabric" node-and-arrow diagram: CCI, UCI, MCP Confidence feeding DLP Agent, Insider Risk, AI Risk; third-party signals via Cloud Exchange; outputs to Real-Time Policy, Analyst Case Surface, AI Command Center.

SCORES AS TIME SERIES: THE TEMPORAL ARCHITECTURE

Trajectory is a scored quantity, not a chart decoration

Most scoring systems are built to identify the entities as put in front of them. The systems described in this paper are built to answer a harder question: "What is this entity becoming?" This distinction is not cosmetic. In several places the derivative of a score is itself a weighted input to another score. The clearest example is the DLP case rubric, where the user's trust trend carries 0.15 of the total weight (half again more than the trust level at 0.10). A user sitting at a mediocre but stable UCI contributes less case risk than a user whose UCI has fallen more than 500 points, because a collapse in trust is evidence in a way a longstanding plateau is not. The Insider Risk agent reads 90 days of UCI history for exactly this trend analysis, and incident velocity (the 30-day count of DLP incidents, tiered at published breakpoints) is a time-bracketed rate, not a lifetime total, so an old burst of activity ages out of the score rather than haunting it.

The AI Risk agent extends the same approach to finer time scales. Session behavior blends a measure of 'burstiness' computed over 12-hour activity buckets, and a single upload spike of a megabyte of data or more surrounded by silence is treated as an exfiltration shape, and scored high on its own. In this example, the temporal pattern, not the byte count, is the signal. Novelty is handled as an explicit decay curve (shown in Figure 3): an application newly observed in the tenant carries the full 100 novelty points during its first week, sliding to 30 by day 30 and 10 by day 90, then zero. New AI surfaces receive scrutiny precisely when they are least understood, and that scrutiny decays on a published schedule instead of at a vendor's discretion. A related seven day trigger adds anomaly weight when a new AI application appears in an individual user's activity.

Scores that move, and the record of their movement

Movement is only trustworthy if it is recorded. A scored DLP case is not frozen at investigation time. When new evidence is delivered, the agent compares the updated evidence profile against a snapshot from the last investigation, and re-investigates when the difference is substantive. This is capped at one re-investigation per case per day, with every score change timestamped in the case history. UCI movement is visible on an event-by-event basis on the user's behavior analytics timeline. Each deduction comes with its source and reason, each third-party ingestion as a searchable anomaly, each revert and each audited reset as its own entry, so the trajectory an analyst sees is assembled from records, not summarized from memory. For MCP, a customer recalculation is explicitly labeled a point-in-time computation and displayed beside the maintained default, which keeps a tenant's yesterday from being mistaken for the catalog's today. This produces the following loop: Telemetry moves scores deterministically, analysts disposition scored results, and dispositions return as labeled outcomes that re-tune the weights. This happens offline, so calibration improves the system without a live feedback loop amplifying it.

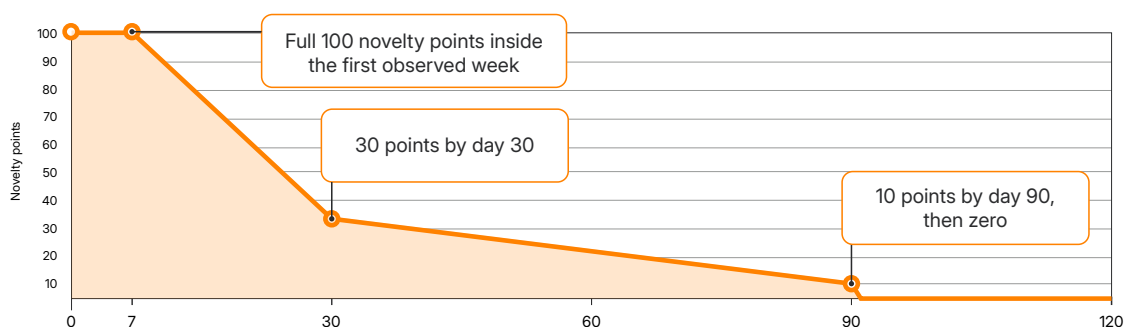


Figure 3: AI Risk novelty decay curve
(100 points in week 1, 30 by day 30, 10 by day 90, then zero).

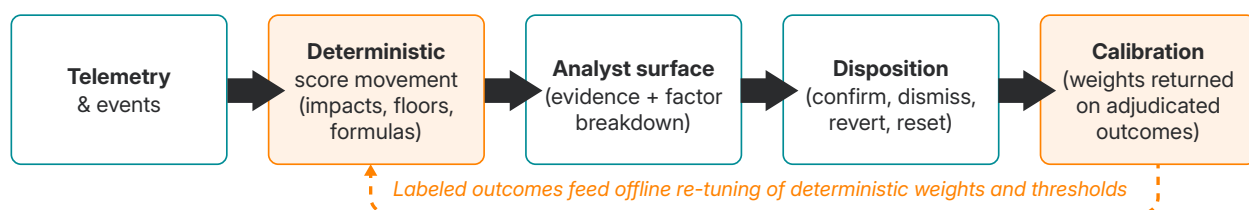
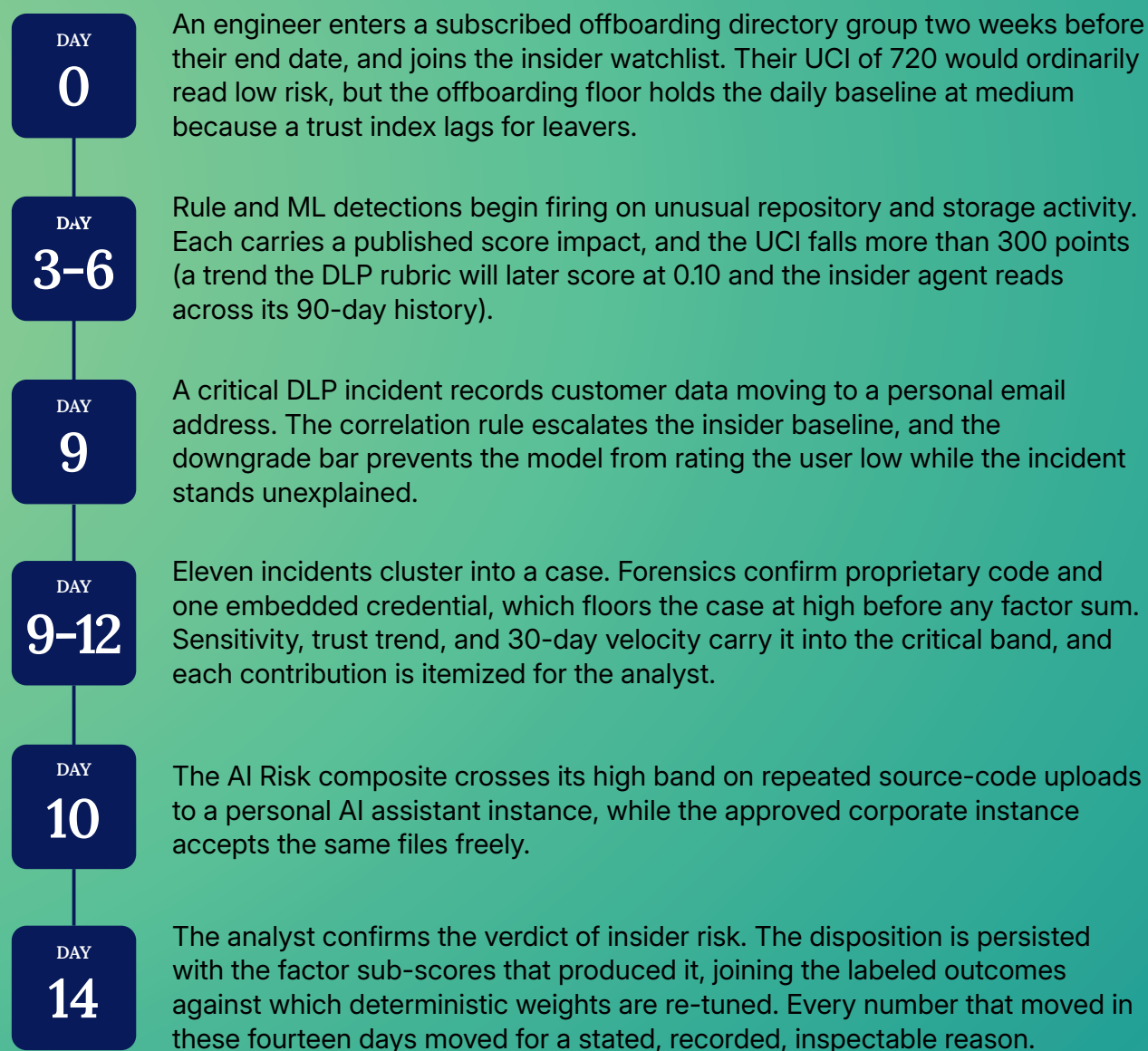


Figure 4 - the score lifecycle loop (Telemetry, Deterministic score movement, Analyst surface, Disposition, Calibration; offline re-tuning path).

In practice: fourteen days of a departing engineer



THE FIVE BUYER QUESTIONS

Carry this table into an evaluation

Industry analysts advise buyers to interrogate any vendor-supplied score on five dimensions, and this evaluation framework is now a regular criteria in RFPs. Table 2 consolidates the preceding sections.

QUESTION	CCI	UCI	MCP CONFIDENCE	AISECOPS
How is it calculated?	Vendor responses plus observed capabilities, numerically combined and normalized 0–100 with category-level rewards and penalties	Published per-severity score impacts (Critical 251–350 down to Informational 1–50) deducted from a 1,000-point baseline	Weighted average over provider-published risk attributes across six families	Published rubrics, weights, and formulas; deterministic floors; AI Risk fully arithmetic
What ground truth?	CSA Cloud Controls Matrix criteria; vendor-supplied and Netskope-observed evidence	Detections from rules and ML models, each visible with source and reason on the user timeline	Provider-published attribute evidence, each with a stated confidence level	Analyst adjudications persisted with sub-scores; thresholds set against production snapshots
How fresh?	Continuous research; on-request assessment of new apps	Event-driven; 90-day anomaly search window; 90-day trend history	Catalog-maintained; customer recalculation is explicitly point-in-time	Daily for Insider and AI Risk; on creation and substantive new evidence for cases
Can it be reweighted?	Yes — per-app sliders across seven categories, with reset (additional license)	Yes — per-policy impacts editable; third-party impact, revert, and reset via REST API	Yes — per-attribute override, recalculation, and reset in-product	Steered by natural-language guidance in agent memory; deterministic weights retuned against adjudicated outcomes
Who set the thresholds?	Netskope research defaults; CCL bands published; tenant weighting adjustable	Severity bands published per policy; admins edit with change review	Confidence bands displayed in the interface (Excellent = 90–100)	Deterministic defaults published in the technical brief; agent behavior steered in natural language within fixed floors

Table 2: The five transparency questions answered per scoring system.

THE NEXT FIVE QUESTIONS (AND NETSKOPE'S CURRENT ANSWERS)

Extending the RFP criteria

The five-question frame is necessary, but not sufficient. A vendor confident in its scoring should be extending the criteria rather than merely surviving them. Netskope proposes five further questions for RFP inclusion, and answers each against its own current state (including where the answer is a boundary rather than a capability). We do this because criteria a vendor writes to flatter itself will not survive contact with a competent evaluator.

1. **Can you show me the score's history?**

A current score without its trajectory hides more than it reveals. UCI exposes its full event-level history on the user timeline, and AISecOps case scores are timestamped through every re-investigation. For CCI and MCP Confidence, the maintained default and any tenant customization are always displayed as distinct values. Entity-level score-history views for these catalogs are the natural next publication step, and buyers should ask every vendor for last-assessed timestamps on catalog scores.

2. **What happens to entities you have not rated?**

The unrated case is the common case in fast-moving categories, and silence about it is a scoring decision made implicitly. Netskope states its defaults, per the "unknown is never treated as safe" principle. An uncataloged MCP server carries a posture score of 80 out of 100 by policy, a missing UCI anchors the insider baseline at medium rather than low, absent telemetry yields inconclusive rather than legitimate, and the DLP rubric redistributes missing factors' weight instead of defaulting downward. Unrated is treated as unknown, and unknown is never treated as safe.

3. **How do you resist gaming of self-reported evidence?**

Two of the four systems consume evidence that an interested party publishes: vendor questionnaire responses in CCI and provider-published attributes in the MCP Catalog. The mitigations are structural. CCI pairs vendor responses with independently observed capabilities, and every MCP attribute carries its own confidence level rather than presenting uniform certainty, so an optimistic claim arrives visibly less trusted than a verified one. In AISecOps, attacker-influenceable content such as filenames and prompt text is treated as evidence the model must cite, not instructions it must obey. Outputs are validated against fixed schemas and allowlisted values, and no quantity of adversarial text can move a score below a deterministic floor.

4. **Who can change the score, and where is the change recorded?**

Every scoring surface in this paper pairs its customization control with a durable record. Policy-impact edits pass through a pending-changes review that shows what changed and who changed it before it applies. UCI resets are written to the incident record as admin-attributed anomalies. MCP attribute overrides are displayed beside the untouched default with a reset path. And CCI risk weights carry an explicit reset to the research baseline. The pattern is uniform: tenant authority over thresholds and weights, vendor custody of the defaults, and an audit trail between them.

5. Do you publish your calibration numbers?

This is the question Netskope believes will decide the category, and the honest current answer is that the instrumentation exists, and publication needs focus. Because AISEcOps persists analyst dispositions alongside factor sub-scores, and because AI Risk thresholds were set against production snapshots with a stated target, the raw material for quantitative calibration reporting (score distributions, band populations, adjudication agreement rates) is retained by design rather than reconstructed after the fact. Buyers should put this question to every vendor whose score they are asked to trust, and should weigh heavily the difference between a vendor that retains labeled outcomes and one that cannot say what its scores were right about.

"Criteria a vendor writes to flatter itself will not survive contact with a competent evaluator."

LIMITATIONS AND HONEST BOUNDARIES

- ML-based UCI detections require additional licensing, so the reweighting answers above assume the relevant entitlements.
- The MCP Catalog's public-server inventory is growing from an initial set and does not yet cover private or self-hosted servers, and its attribute evidence is only as good as what providers publish. This is why each attribute carries its own confidence level, rather than presenting uniform certainty.
- Two AI Risk categories that would require per-server data attribution and tool-scope manifests are deliberately left out of the MCP composite until those signals exist upstream; the scoring model does not simulate inputs it does not have.
- AISEcOps scores are advisory by design: the agents rank and provide evidence, and disposition authority remains with the security team.
- No score in this portfolio observes traffic that never crosses a Netskope enforcement point.

These boundaries are stated because a scoring system that cannot articulate what it does not know invites exactly the skepticism this paper exists to answer.

SCORES ARE JUDGMENT MADE LEGIBLE

A platform's scores are its judgment made legible, and the quality of that judgment is now a procurement criterion in its own right. This paper has argued that Netskope's answer is structural rather than rhetorical at three levels.

At the level of principle: One question per scale, deterministic floors beneath every model, evidence with every number, tenant dials over research defaults, durable records of movement, and honest treatment of the unknown.

At the level of history: Four eras of expansion, each adding a scorer for a new entity class under the same grammar rather than a new grammar per product.

At the level of the fabric: One-way signal paths, explicitly cut loops, audited openness to third-party judgment, and scores that exit as enforceable policy and consistent labels.

Each claim is inspectable in product documentation or in a licensed tenant.

Organizations drafting RFP criteria for AI risk scoring are invited to use Tables 1 and 2 and the five questions of Section 11 as the template, and to hold every score they are offered - including Netskope's - to its light.

REFERENCES

¹ Netskope Knowledge Portal, "[App Catalog](#)" (Cloud Confidence Index methodology and Risk Weight Settings).

² Netskope Knowledge Portal, "[Cloud Confidence Index](#)."

³ Cloud Security Alliance, "[Cloud Controls Matrix \(CCM\)](#)."

⁴ Netskope Knowledge Portal, "[ML Based Policies](#)."

⁵ Netskope Knowledge Portal, "[Using the REST API v2 UCI Impact Endpoints](#)."

⁶ Netskope Knowledge Portal, "[Using the REST API v2 UCI Reset](#)."

⁷ Netskope Knowledge Portal, "[App Catalog and Risk Assessment](#)" (MCP Catalog).

⁸ Netskope, "Risk Scoring in Netskope AI SecOps," Technical Brief, July 2026.

⁹ Netskope Knowledge Portal, "[AI Command Center](#)."

Interested in learning more?

[Request a demo](#)



Netskope, a leader in modern security and networking, addresses the needs of both security and networking teams by providing optimized access and real-time, context-based security for people, devices, and data anywhere they go. Thousands of customers, including more than 30 of the Fortune 100, trust the Netskope One platform, its Zero Trust Engine, and its powerful NewEdge network to reduce risk and gain full visibility and control over cloud, AI, SaaS, web, and private applications—providing security and accelerating performance without trade-offs.

©2026 Netskope, Inc. All rights reserved. Netskope, NewEdge, SkopeAI, and the stylized "N" logo are registered trademarks of Netskope, Inc. Netskope Active, Netskope Cloud XD, Netskope Discovery, Cloud Confidence Index, and SkopeSights are trademarks of Netskope, Inc. All other trademarks included are trademarks of their respective owners. 09/26 WP-1025-1