PITON Global

EXECUTIVE WHITE PAPER · WP-54

CONTENT INTEGRITY · PHILIPPINES

The integrity-signal standard: the executive economics of content integrity outsourcing to the Philippines.

An analysis of why signals reviewed is a volume vanity metric, how threat-catch accuracy and platform trust — never integrity throughput — decide the true cost of a content-integrity operation once missed coordinated abuse, false enforcement, escalation misses, and eroded trust are counted, and the vendor-selection discipline that catches the real threat and leaves legitimate users alone.

AUTHORS

John Maczynski — Chief Executive Officer
Ralf Ellspermann — Chief Strategy Officer

JULY 2026
MANILA · ESTABLISHED 2001

Contents

ABOUT THIS SERIES — VOLUME 54

Every content-integrity proposal is priced per signal reviewed, because volume is easy to count. This volume is about the only integrity outcome a platform's trust actually depends on: the real threat caught — the coordinated abuse, the fraud ring, the manipulation campaign — and enforced precisely, without sweeping up legitimate users. The gap between reviewing a signal and catching the threat behind it is where missed abuse and eroded trust live. The full series lives at piton-global.com.

THE PITON-GLOBAL LEADERSHIP TEAM

About the authors

John Maczynski

Chief Executive Officer

John has bought platform-integrity capacity for three decades, and learned that the fast operation that reviews every signal and misses the coordinated campaign is the most expensive one a platform can hire — because a missed abuse network is a trust crisis in waiting, and a false enforcement sweep is a wave of wrongly-punished users and a press cycle. His standing question ignores the signals-cleared report: of the signals you reviewed, how many real threats did you catch, and how precisely did you enforce? Applied at PITON-Global, that question is what clients get on every integrity sourcing decision — personally, with no account layer in between.

j.maczynski@piton-global.com
+1 402-598-8740

Ralf Ellspermann

Chief Strategy Officer

Ralf has spent twenty-five years running Philippine platform-integrity floors, and he judges an integrity team by its threat-catch accuracy and enforcement precision, not its signals-per-hour: whether the real abuse was caught and enforcement landed only on it, or the campaign slipped through while legitimate users got swept up. The integrity contract in Section 3 codifies what he learned rebuilding teams that cleared signal queues fast and missed the threats. He now reads integrity providers the way he once ran them — from the missed-threat log and the false-enforcement rate backward.

r.ellspermann@piton-global.com
Manila, Philippines

PITON Global

ABOUT PITON-GLOBAL

PITON-Global is a buyer-side advisory practice, Manila-based since 2001, that has never operated a delivery floor or taken a placement commission. Its stock-in-trade is knowing the Philippine provider market at the operating level — 1,000+ mapped operations, re-verified continuously, including each integrity team's real threat-catch accuracy and enforcement precision rather than its signals-per-hour

headline — and converting that knowledge into a shortlist of six to ten genuinely qualified partners per engagement. Buyers from Tripadvisor, Garmin, TSYS, HealthPartners, Yetipay, Workrise, and Norman Disney & Young have run their searches through the practice.

SECTION 01

Executive summary

-56%
Fully-loaded cost per accurate integrity action vs. US onshore

97%+
Threat-catch accuracy on vetted teams, up from 78–88%

-67%
Reduction in false enforcement on legitimate users, vetted teams

6.2×
First-year ROI, reconstructed case (Section 6)

A content-integrity engagement is bought on the price per signal and paid for on the threats it misses and the users it wrongly punishes. The vendor reports millions of signals reviewed at a fraction of the onshore cost; the platform finds the coordinated inauthentic network that operated for months undetected, the fraud ring that scaled while its signals sat in a cleared queue, the false-positive enforcement that banned real users and drew backlash, and the integrity metrics drifting while activity looked healthy. The operation cleared its signal queues and the platform inherited an abuse problem, a false-enforcement backlash, and eroded trust. The gap between reviewing a signal and catching the threat behind it is where the real cost of cheap integrity work hides, because a missed threat or a false enforcement is not throughput delivered — it is abuse shipped or a legitimate user harmed.

The Philippines — English-fluent, culturally fluent with Western platform norms, and with a mature platform-integrity talent pool — is where content integrity is delivered with the most threat-catch and precision discipline, because that pool can run the pattern analysis, precise enforcement, and judgment-grade escalation that integrity work demands. Five findings:

1 **Signals reviewed is the industry's most flattering number.** Any floor can clear a signal queue at speed. Vetted teams are measured on outcome: threat-catch accuracy, enforcement precision, and escalation quality. Weak providers quote a low per-signal price the platform repays in missed abuse, false enforcement, and lost trust. Section 2 shows the gap.

2 **The caught threat and precise action, not the fast clear, is the product.** Integrity that protects a platform comes from seeing the pattern behind the signals, enforcing only on the real threat, and escalating the

novel case — not from clearing a queue. Teams that catch threats protect trust; teams that clear loosely miss abuse and punish the innocent, at a cost far above the per-signal price.

3 **An integrity miss cuts both ways — threat missed and user wronged.** Miss the coordinated campaign and abuse scales unchecked; over-enforce and you ban legitimate users and invite backlash and regulatory scrutiny. Vetted teams hold threat-catch above 97% and false enforcement low precisely because both tails cost — the abuse that slips through and the user wrongly punished.

4 **The missed campaign and the backlash, not the signal, are what cost.** A threat missed costs the abuse it enables, the incident response, and the trust it erodes; a false enforcement costs the appeals wave and the reputational hit — many times the per-signal saving. Teams that catch threats and enforce precisely protect trust and revenue, and that protected trust — not the offshore rate alone — is where the case study's return is built.

5 **Contract on the accurate integrity action, not the reviewed signal.** The contracting failure of integrity work is paying per signal and hoping the threats are caught and enforcement is precise. Vetted deals price against threat-catch floors, false-enforcement ceilings, and escalation-quality targets, making accurate integrity outcomes the team's problem, which is where it belongs.

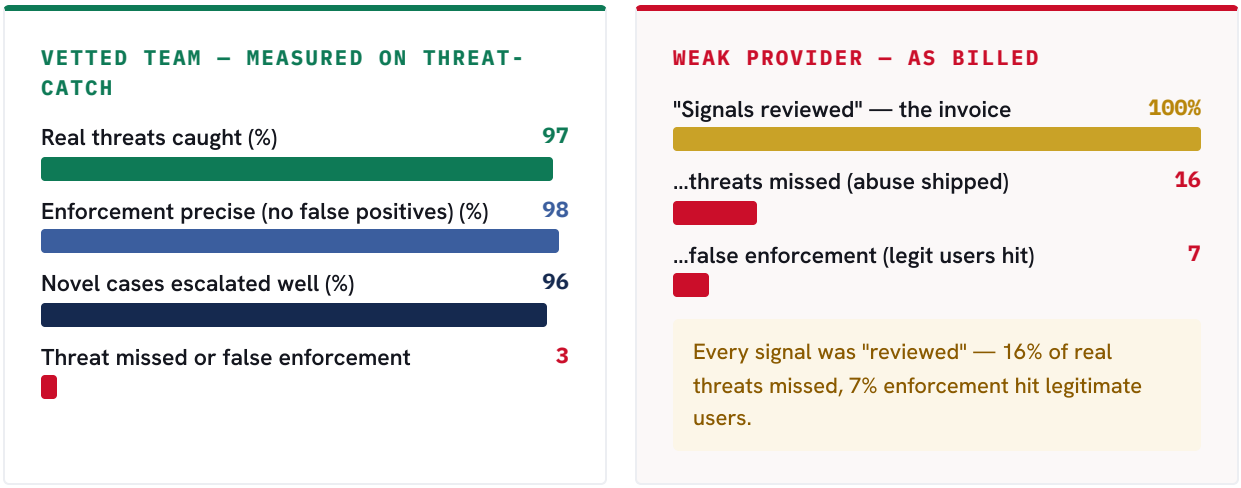
The intended reader is a head of platform integrity, trust-and-safety leader, or COO buying content integrity per signal and paying for it in missed threats, false enforcement, and eroded trust. Sections 2–4 separate the teams that catch threats from the teams that clear signals; Sections 5–7, the method and the proof.

SECTION 02

The volume mirage: signals reviewed versus threats caught

Figure 1 shows one month of a platform's integrity queue two ways: as the signals-reviewed report a weak vendor bills, and as the threat-catch reality the abuse landscape and the appeals queue experience. The signals-reviewed number is identical in spirit to a chatbot's containment rate — a figure engineered to look like output while missed threats, false enforcement, and eroded trust accumulate underneath.

FIG. 1 — ONE MONTH OF SIGNALS: THREAT-CATCH QUALITY VS. REPORTED VOLUME



Left: PITON-Global engagement instrumentation, 2025–26, threat-catch audited via red-team and backtest. Right: composite of pre-engagement vendor reporting reviewed in diligence.

The economics follow the threat-catch and precision, not the signals count. On the left, ninety-seven percent of real threats are caught and enforcement lands only on them, so abuse is contained and legitimate users are left alone; on the right, a cheap per-signal price misses sixteen percent of threats and wrongly enforces on seven, paying in shipped abuse, backlash, and eroded trust at once. A low per-signal price that misses threats and over-enforces is not a saving; it is abuse exposure and trust erosion, delayed and compounding. The threat-catch-and-precision audit — catch accuracy, false-enforcement rate, escalation quality — is the cheapest due-diligence instrument in this paper.

SECTION 03

The integrity contract: catch the real threat, enforce precisely, escalate what matters

Figure 2 lays out the three clauses of a working content-integrity contract — the machinery a floor read verifies before any per-signal price is believed.

FIG. 2 — THE INTEGRITY CONTRACT, AS OPERATED ON VETTED TEAMS

CATCH THE REAL THREAT

Signals are analyzed for the pattern behind them — coordinated inauthentic behavior, fraud rings, manipulation campaigns — not triaged one at a time, so the network is seen and caught rather than each node cleared in isolation. The threat is found where it actually lives, in the connections between signals.

ENFORCE PRECISELY

Enforcement lands on the actual bad actors and their assets, scoped to the policy and the evidence, with legitimate users left untouched and a review step on high-impact actions — so abuse is removed without a false-positive sweep. Precision is what separates integrity from collateral damage.

ESCALATE WHAT MATTERS

Novel attack patterns, emerging campaigns, and policy-gap cases are escalated to the platform's integrity team with evidence and context, feeding detection back into the system — with a threat-and-escalation report the platform can read. Escalation is how a floor stops being reactive and starts closing the loop on new abuse.

↓ yields ↓

THE ACCURATE INTEGRITY ACTION

97%+ THREAT-CATCH | PRECISE ENFORCEMENT | LOOP CLOSED

Real threats are caught, enforcement lands only on bad actors, and novel abuse is escalated back into detection — so coordinated abuse is contained, legitimate users are protected, and platform trust holds. The operation gets cheaper per accurate integrity action and safer at the same time, which is the entire promise of outsourcing integrity to people who measure threat-catch and precision, not signals reviewed.

Operating parameters from PITON-Global platform-integrity engagements, 2025–26. Integrity analysts work alongside client detection systems and escalate to in-house teams.

In diligence, the contract is tested from the threat's side: run a red-team and backtest a known-abuse dataset to measure true threat-catch and false-enforcement rates; read the escalation log for quality; interview on how they detect a coordinated campaign versus organic behavior. Perhaps one team in ten survives it cleanly. Steps 3 and 5 of the framework exist to find that one before the contract does — because everything in Table 3 of Section 6 depends on threat-catch accuracy, not signals reviewed.

SECTION 04

Cost and performance: cost-per-accurate-

integrity-action economics and benchmarks

TABLE 1 — FULLY-LOADED COST PER ACCURATE INTEGRITY ACTION: THREE OPERATING MODELS (2026)

OPERATING MODEL	COST / ACTION	VS. US BASELINE
US onshore in-house integrity team	$1.62	baseline
Low-cost offshore signal mill (per-signal)	$1.12	−31%
Vetted Manila team — threat-catching, precise	$0.71	−56%

Basis: PITON-Global engagement pricing 2025–26; cost per accurate integrity action = fully-loaded operating cost divided by actions that correctly caught a real threat or correctly cleared a signal (missed threats and false enforcement excluded); the signal mill looks cheap per signal but missed abuse, false-enforcement backlash, and incident response lift its true cost per accurate action; US comparator fully-loaded in-house cost.

TABLE 2 — CONTENT-INTEGRITY BENCHMARKS, VETTED VS. BASELINE (2025–26)

METRIC	VETTED TOP TIER	BASELINE	EXECUTIVE READING
Threat-catch accuracy (recall)	97%+	78–88%	The only integrity number trust follows
Enforcement precision	98%+	85–92%	Bad actors only, or collateral damage
False enforcement vs. baseline	−67%	baseline	"Enforce precisely" — held or not
Coordinated-abuse detection time	−58%	baseline	Caught early, or scaled unchecked
Escalation quality (actioned by platform)	94%+	70–82%	Signal to the core team, or noise

Source: PITON-Global platform-integrity engagement data 2025–26; baseline = typical pre-engagement operations billed on signals reviewed.

The strategic read: the signal mill's −31% per signal is a false economy — missed abuse, false-enforcement backlash, and incident response lift its true cost per accurate action far above the per-signal gap, and trust lost to an undetected campaign is a cost no rate recovers. The vetted Manila team's −56% is real because it is measured where the value is: the real threat caught, enforcement precise, the novel case escalated. Buy

on signals reviewed and you buy the right column of Figure 1; buy on cost-per-accurate-integrity-action and you buy threat-catch accuracy — the same discipline every volume in this series keeps arriving at from a different door.

SECTION 05

Sourcing discipline: the 7-Step Vendor Vetting Framework, applied to content integrity

Content-integrity diligence adds one hard requirement to every step: a per-signal price is real only when the threat-catch accuracy behind it survives a red-team and backtest against a known-abuse dataset. The funnel below is representative of a 60–150-seat platform-integrity search.

FIG. 3 — THE 7-STEP FRAMEWORK AS A CONTENT-INTEGRITY FUNNEL

	STEP	WHAT IT ELIMINATES IN AN INTEGRITY SEARCH	FIELD REMAINING
1	**Requirements definition**	Throughput-first scoping: fixes the threat-catch floor, the enforcement-precision target, the false-enforcement ceiling, and the escalation-quality standard the operation must meet by abuse type	1,000+
2	**Market mapping**	Signal mills: operations that bill reviewed volume and treat missed threats and false enforcement as the client's problem, and generalist floors with no threat-analysis or investigations depth	~46
3	**Capability screening**	Teams without the machinery: audited threat-catch history, pattern-analysis and investigations capability, a red-team/backtest QA loop, and coverage across your abuse vectors and languages	~13
4	**Compliance & security audit**	Weak governance: user-data protection, secure-facility controls, regulatory alignment, insider-threat and access controls on integrity tooling, SOC 2 scope covering the integrity stack	~8
5	**Operational diligence**	The red-team/backtest: run a known-abuse dataset for true threat-catch and false-enforcement; read the escalation log for quality; interview on how they detect coordinated vs. organic behavior	4–6
6	**Competitive RFP & negotiation**	Per-signal pricing: finalists bid against threat-catch floors, enforcement-precision targets, and false-enforcement ceilings — all carrying credits for missed threats or wrongful enforcement	2–3
7	**Transition & launch governance**	Cold start: the team calibrates against your abuse taxonomy, detection tooling, and escalation paths on a backtest before it owns live signals, with the threat-catch floor certified before steady state, PITON-Global embedded buyer-side	1

Field-remaining counts represent a 60–150-seat platform-integrity search, PITON-Global engagements 2024–26.

The standing guarantees hold: **vendor neutrality** — PITON-Global operates no integrity floors, actions no signals, and takes no placement economics, so the shortlist reflects audited threat-catch and precision performance, not inventory; **right-sizing** — teams for whom your operation is a flagship account. Six to ten

genuinely qualified providers, delivered within 48–72 hours, free to the buyer, competing through a fully transparent, fair, comprehensive, and highly competitive RFP.

SECTION 06 · ANONYMIZED CASE STUDY

Re-basing the operation on threat-catch accuracy: a 90-seat platform-integrity rebuild, reconstructed

Engagement CI-054 is a US digital platform — processing roughly 70,000,000 integrity signals a year across spam, fraud, and coordinated-behavior vectors — that had offshored content integrity to a per-signal mill. The signal price was excellent; the threat-catch was not: catch accuracy sat near 84%, coordinated campaigns were operating undetected for weeks, false enforcement was drawing user backlash, and escalations were mostly noise. The head of platform integrity wanted the scale without the missed abuse and the collateral damage. Identifying details are anonymized under NDA.

SELECTION (WEEKS 0–8, RED-TEAM/BACKTEST)

Requirements fixed a 97% threat-catch floor, a 98% enforcement-precision target, and a false-enforcement ceiling, with a mandatory pattern-analysis and escalation workflow. Mapping produced 46 claimants; the machinery screen — threat-catch history, investigations capability, red-team/backtest loop, vector coverage — left 13; the security-and-insider-threat audit left 8. Blind red-team and backtest runs against a known-abuse dataset shortlisted 5. A Manila team won priced against threat-catch and precision, with a backtest written into the reporting spec.

TRANSITION (MONTHS 2–7, CALIBRATED FIRST)

The team calibrated against the platform's abuse taxonomy, detection tooling, and escalation paths on backtests before owning live signals; the threat-catch floor was certified in month 3. The red-team-and-feedback loop fed detection back continuously. By month 7 threat-catch reached 97.3%, enforcement precision hit 98%, false enforcement fell 67%, and coordinated-abuse detection time dropped 58%. PITON-Global chaired the threat-catch review board to steady state.

TABLE 3 — CI-054 FIRST-YEAR VALUE BUILD-UP (US$)

VALUE COMPONENT	YEAR 1	BASIS
Labor arbitrage on the 90-seat operation	**$2.36M**	168.8k productive hrs × ($23.90 − $9.90)
Abuse & incident-response cost avoided	**$0.84M**	Undetected-campaign and IR cost removed
Trust & false-enforcement backlash avoided	**$0.44M**	Appeals wave and reputational exposure removed
Rework & re-review reduction	**$0.18M**	Re-review of wrong actions removed
Gross value created	**$3.82M**	
Less: engagement & transition costs	**($0.62M)**	Calibration, backtest build, tooling/escalation integration; advisory fee $0 (provider-paid model)
Net value ÷ cost = first-year ROI	**6.2×**	$3.82M ÷ $0.62M

Figures rounded; catch, precision, and detection-time effects audit-verified. Method in Methodology & notes.

Sixty-two percent arbitrage, thirty-eight percent from avoided abuse and incident response, avoided backlash, and less re-review — a typical split for the series, earned because the operation was re-based on threat-catch accuracy rather than signal volume. Every line traces to a diligence step: the threat-catch floor to Step 1, the catch-and-precision pricing to Step 6, the red-team/backtest loop to Step 3's contract. The head of platform integrity's line at the year-one review: "The signal price barely moved. What moved is that we catch the campaigns now — and we stopped banning our own users."

SECTION 07

The executive playbook: governance for the first 180 days

A content-integrity transition succeeds when the team is measured and paid on threat-catch accuracy and enforcement precision — client in command throughout.

DAYS 0–30

Define threat-catch, contract on it

Set the threat-catch floor, the enforcement-precision target, and the false-enforcement and escalation-quality thresholds with platform-integrity leadership before launch, and document the abuse taxonomy and escalation paths. Contract

	against threat-catch and precision — never per signal alone — with floors and ceilings carrying credits. Stand up the red-team/backtest as the program's arbiter of truth.
DAYS 30–90 **Calibrate, prove the floor**	Calibrate the team against your abuse taxonomy, detection tooling, and escalation paths on backtests before it owns live signals; certify the threat-catch floor and red-team loop. Switch on the escalation-and-feedback loop from day one. Publish the integrity dashboard: threat-catch, enforcement precision, false-enforcement rate, detection time, escalation quality.
DAYS 90–180 **Own live signals, certify steady state**	Hand live signals to the team only after it holds the threat-catch floor and precision target. Review the integrity board monthly with the missed-threat and false-enforcement trends on the table. Certify steady state on two months at floor including a novel-attack cycle; hand over on documented criteria; keep the quarterly red-team/backtest standing.

THE FIVE FAILURE MODES WE ARE HIRED TO PREVENT

Paying per signal while real threats are missed · clearing signal queues instead of catching the pattern · over-enforcing and sweeping up legitimate users · escalating noise instead of the novel threat · owning live signals before threat-catch is proven. All five are settled at selection and contract.

The content-integrity argument, in one sentence: a reviewed signal is activity and an accurate integrity action is a real threat caught and a legitimate user left alone, and a vetted Philippine team delivers –56% cost per accurate integrity action with 97%+ threat-catch, 98%+ enforcement precision, and false enforcement cut by two-thirds — for buyers who red-team the threat-catch instead of counting signals reviewed, and compete the finalists through a fully transparent, fair, comprehensive, and highly competitive RFP. Anyone can review a signal. We make sure you buy the team that catches the threat behind it.

METHODOLOGY & NOTES

How the numbers were built

Threat-catch and precision figures (Fig. 1, Table 2) derive from engagement instrumentation in PITON-Global platform-integrity engagements, 2025–26. Threat-catch accuracy and enforcement precision are measured by red-team and backtest against a known-abuse dataset; false-enforcement rate, detection time, and escalation quality are measured against the client's pre-engagement baseline. The weak-provider column is a composite of pre-engagement vendor reporting reviewed in diligence, shown for contrast.

Cost-per-accurate-integrity-action figures (Table 1) are fully loaded — wages, benefits and statutory charges, facilities, technology and detection tooling, management and QA overhead including threat analysts and red-team leads, attrition-driven recruiting — divided by actions that correctly caught a real threat or correctly cleared a signal, with missed threats and false enforcement excluded from the denominator; US comparators reflect fully-loaded in-house cost on the same basis. An accurate integrity action catches a real threat, or clears a benign signal, without collateral damage.

The case study (CI-054) is anonymized under NDA; volumes and dates are generalized, figures rounded. Catch, precision, and detection-time effects are audit-verified; the abuse-avoided line reflects incident-response and undetected-campaign cost removed, and the backlash-avoided line is a conservative estimate. ROI = net first-year value ÷ all engagement-related client costs. PITON-Global's advisory fee is provider-paid and appears at $0 in the client cost base; neutrality safeguards are documented at piton-global.com.

Market context draws on IBPAP reporting and PITON-Global's proprietary provider mapping (1,000+ operations; threat-catch accuracy and enforcement precision re-verified May 2026). Integrity analysts work alongside client detection systems and escalate to in-house teams; wellness provisions for exposure-heavy work reflect on-staff support and rotation. Estimates and projections are PITON-Global analysis and should be cited as such. The full series is available at piton-global.com.

Would your provider's signal price survive a red-team backtest?

Thirty minutes with John will tell you. Free to buyers, strictly vendor-neutral — a threat-catch-verified shortlist within two to three business days.

j.maczynski@piton-global.com
+1 402-598-8740