cognizant

From reaction to reinvention:

Redefining trust and safety in gaming

A global study on gaming
trust and safety by Everest Group,
supported by Cognizant

Kanti Kopalle

Vice President, Intuitive Operations
and Automation

Foreword

The gaming industry is no longer just about entertainment - it's a dynamic digital ecosystem where billions of players connect, compete, and create. With over 3.5 billion active gamers globally and mobile gaming contributing more than $126 billion in annual revenue, the pressure to keep these digital worlds safe, inclusive, and resilient has never been higher.

This Everest Group whitepaper, presented by Cognizant, explores the evolving landscape of Trust and Safety (T&S) in gaming. It highlights how enterprises are navigating escalating threats - from harassment and misinformation to AI-generated deepfakes - while facing tighter budgets and growing regulatory scrutiny. Notably, **~90%** of gaming companies have maintained flat or modest T&S budgets, even as risks intensify.

The paper also examines the dual impact of generative AI. While 36% of developers are already using these tools to unlock new creative frontiers, this technology also presents new and complex safety challenges. Striking the right balance between innovation and responsibility is now a critical mandate.

In this landscape, a fundamental shift is occurring. Yesterday's view of Trust and Safety as a mere cost center is dissolving. Today, it is rightfully being recognized as the very bedrock of a thriving digital community - the invisible architecture of player loyalty and platform integrity.

This paper offers a pragmatic roadmap for gaming enterprises, blending Safety by Design, AI-human hybrid moderation, and strategic outsourcing to scale responsibly and reduce the Total Cost of Ownership. It reflects a broader industry evolution: from reactive enforcement to a proactive, integrated, and innovation-led safety ecosystem.

We hope this paper serves as a valuable guide for leaders to reimagine T&S not as a compliance checkbox, but as a powerful catalyst for trust, engagement, and long-term growth.

Everest Group®

August, 2025

From Reaction to Reinvention: Redefining Trust and Safety (T&S) in Gaming

cognizant

Contents

Introduction

The ever-expanding gaming industry has evolved into a leading digital space for social connection, self-expression, and commerce, where Player Experience (PX) determines platform success and longevity. In this environment, enterprises are under increasing pressure to strengthen their T&S capabilities, a foundational pillar of PX to build long-term trust, community loyalty, and enhance monetization.

This necessitates that gaming enterprises align their T&S investments with the growing risk landscape, rising cross-border interactions, and increasing presence of vulnerable player groups, while remaining compliant with global regulatory frameworks. However, even as these risks intensify, many enterprises face tighter budget constraints driven by cautious investor sentiment and tightening monetization channels. In 2024, private funding for the global gaming sector declined by 12% compared to 2023.[1] This contraction reflects a broader downward trajectory that has continued since the industry's investment peak in 2021. Even our survey reveals that about ~90% of enterprises have either maintained flat T&S budgets or reported only a modest increase in the last 12-18 months, underscoring the imperative for more scalable, resource-efficient approaches to safeguarding player safety.

Many enterprises acknowledge that in-house capabilities alone are insufficient to address the scale, complexity, and evolving nature of T&S challenges. Concerns around data privacy, content sensitivity, shrinking budgets, and perceived loss of control have historically prevented enterprises from outsourcing human moderation at scale. Strategic outsourcing, however, can add significant value by enabling operational scale, offering access to specialized expertise, and enabling operational efficiencies, reducing the Total Cost of Ownership (TCO).

This Viewpoint explores how gaming companies can combat declining budgets, rising operational costs, and increasing regulatory compliance requirements through a pragmatic, impact-oriented approach to T&S that leverages the best of both internal capabilities and strategic outsourcing to deliver a safe online platform and superior PX. Drawing on insights from a survey of 50 global gaming enterprises, this Viewpoint delves into several key themes in gaming, including:

- An overview of evolving global T&S priorities of different gaming enterprises
- The double-edged impact of gen AI on the gaming experience
- In-house measures enterprises adopt to mitigate risks
- Understanding why enterprises hesitate to outsource despite its benefits
- How outsourcing can be leveraged as a strategic lever to reduce TCO
- How to optimally leverage outsourcing to complement internal capabilities
- The future of T&S in gaming

[1] Why are gaming start-ups struggling to attract funding in 2025, GREY Journal

Gaming companies’ T&S imperatives

The need for robust T&S measures

Gaming enterprises face a diverse and evolving range of threats that span player safety and wellness (for example, harassment, harmful content, and account security), community health (extremism, child exploitation, and misinformation), legal and regulatory compliance (age verification, data privacy), game integrity (fraud, cheating, and IP infringement), AI safety risks (fairness, explainability, misuse), and operational risks (technical vulnerabilities, system abuse, moderation backlogs).

Exhibit 1 highlights the core risk categories, as chosen by respondents, driving their T&S spend.

Exhibit 1: T&S threat vectors driving enterprises’ T&S spend

Source: Everest Group (2025)

89% **Account compromise or fraud** (including phishing, credential theft, stolen items/currency)

84% **Hate speech and toxic behavior** (including hate speech and in-game bullying)

84% **Cheating and unfair gameplay** (including bots, hacks, wallhacks, and exploitative rank climbing)

76% **Scams and social engineering** (including fake giveaways, impersonation, marketplace fraud, and deepfakes)

58% **Child safety risks** (including grooming, inappropriate interactions, and data privacy violations)

42% **Misinformation or disruptive content** (including manipulation, misleading in-game chat, and deceptive gameplay tactics)

37% **IP theft or unauthorized content sharing** (including pirated mods and unlicensed use of assets)

While some of these threats are universally critical, the relative emphasis varies across gaming segments. Exhibit 2 lists the T&S challenges different gaming segments face.

Exhibit 2: T&S challenges different gaming segments face

Source: Everest Group (2025)

Battle royale and tactical shooter games (such as Fortnite, Call of Duty, Warzone)	Addressing **real-time** abuse in voice and chat during high-intensity gameplay, including harassment, slurs, and rage-related behavior Combating **cheating methods**, such as aimbots, wallhacks, and smurfing, that undermine fairness and trust Mitigating **anonymity-based harassment** and ensuring rapid responses to live incidents
Competitive arenas and esport titles (such as League of Legends, Valorant, Dota 2)	Preventing **coordinated abuse** and harassment in ranked matchmaking and tournament settings Ensuring **consistent policy enforcemen**t to uphold fairness across formal and informal esport environments
UGC and sandbox environments (such as Roblox, Minecraft, Garry's Mod)	**Moderating UGC at scale**, including assets, avatars, and in-world experiences that may be exploitative or harmful Preventing **real-time abuse** and **platform data exposure** due to unfiltered user-created environments
Real-money and betting platforms (such as Stake.com, PokerStars, Skillz)	Enforcing **age and identity verification** to comply with gambling regulations and prevent underage access Detecting and preventing **match manipulation and financial fraud** linked to betting outcomes **Social engineering** for account hijacks or fund redirection
Immersive virtual worlds and metaverse platforms (such as Decentraland, Second Life, The Sandbox)	Enforcing **age verification and child privacy controls** to protect underage users and comply with global regulations Mitigating **platform data leakage and real-time exposure risks** through access controls and secure infrastructure Safeguarding users from **grooming and exploitation** through avatar/proximity-based interactions
Social party games (such as Among Us, Jackbox Party Pack, Fall Guys)	Combating **impersonation** and **trolling** in anonymous or nickname-based lobbies Addressing **grooming, inappropriate content, and harassment** in light-moderation environments
Action RPGs and co-op exploration games (such as Diablo IV)	Preventing **phishing scams** through impersonation of official support or player accounts Tackling **griefing behaviors** in co-op or matchmaking sessions that disrupt gameplay
Multi-media RPGs and role-play-based games (such as Cyberpunk 2077, Elden Ring)	Addressing **scams, phishing, and fraudulent trading** practices targeting in-game currencies and virtual goods Mitigating **account takeovers and impersonation** attempts exploiting player trust in persistent communities

Robust T&S measures enable enterprises to build safer and high-trust gaming environments, reducing exposure to harmful content and reinforcing confidence in the platform's ability to protect its community. 20% of players are likely to spend less money in gaming spaces due to harassment.[2] Gaming enterprises are also increasingly using positive reinforcement mechanisms to foster a more welcoming player environment, which drives higher retention and organic growth. For example, Riot Games' revamped Honor system rewards positive player behavior with in-game incentives, leading to improved conduct and higher player retention. Further, as enterprises adopt AI for personalization, content creation, and moderation, robust T&S ensures that innovations remain user-friendly, transparent, and responsible for enhancing PX without compromising fairness, inclusion, or safety.

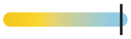
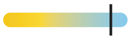
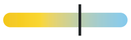
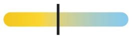
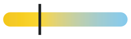
Gaming enterprises' T&S investment priorities

In response to the T&S challenges, gaming enterprises are aligning their investment priorities not only to manage current threats but also to future-proof against emerging threats, regulatory shifts, and player trust erosion. Exhibit 3 captures insights from our survey on key investment priorities in T&S for gaming enterprises.

Exhibit 3: Gaming enterprises' T&S investment priorities

Source: Everest Group (2025)

Enterprise relatability Low → High

Detecting and preventing cheating, scams, financial fraud, and unauthorized account access 	Real-time detection of phishing attempts, scams, and credential stuffing	Integration of anti-cheat engines and fraud scoring systems across user accounts and transactions	
Reviewing and moderating content 	Scalable moderation of offensive content (including AI-generated avatars, mods, maps, NFTs)	Strengthening moderation pipelines for multi-language and multi-platform deployment	Balancing automation with human review for nuanced edge cases
Developing Safety by Design principles 	User journey design to minimize exposure to risk	Embedding safety input in early-stage product prototyping	Safety-led design reviews
Ensuring compliance with global digital safety regulations 	Aligning platform policies with regulations such as GDPR, COPPA, and the EU DSA	Implementing user consent workflows and privacy disclosures for AI-driven moderation	Educating players about regulatory standards and platform safety commitments
Implementing child safety measures 	Robust age verification and guardian consent systems	Default high-privacy settings for underage users	Preventing grooming via chat/voice

These evolving T&S priorities are set to take on even greater urgency with accelerating gen AI adoption across the gaming ecosystem.

[2] Study reveals the impact of online harassment in games on player spending, GoNintendo

Impact of gen AI on T&S in gaming

Gen AI is opening a transformative frontier for the gaming industry. From automating world-building and character design to enriching narrative depth and accelerating Quality Assurance (QA) processes, gen AI is reshaping every phase of the game development life cycle. Its influence extends well beyond creation, driving innovation across marketing, live operations, community management, and player support, introducing new levels of personalization, efficiency, and scalability. Exhibit 4 highlights the key areas in which gaming enterprises leverage gen AI for gaming operations.

Exhibit 4: Top gen AI applications in gaming

Source: Everest Group (2025)

63%

Automating QA and feature testing using gen AI for **bug detection** and **balance checks**

61%

Using gen AI to **generate code** or **optimize gameplay features** such as scripts

58%

Using gen AI for **narrative design and story building**

55%

Creating in-game **assets with gen AI**, including characters, maps, mods, and skins

55%

Using gen AI for **drafting community engagement** content such as posts, patch notes, and player communication

50%

Using gen AI for **dialogue scripting** or NPC interaction designing

However, gen AI also amplifies existing T&S risks and introduces new, complex vectors of harm such as AI-generated voice and visual deepfakes, stylized toxicity, and synthetic spam. Its capacity to autonomously generate assets and real-time in-game content presents significant challenges related to IP ownership, copyright liability, and the lawful use of training data. Exhibit 5 captures insights from our survey on how gen AI has influenced the existing T&S threat landscape.

Exhibit 5: Impact of gen AI on existing T&S threat landscape

Source: Everest Group (2025)

55%

Gen AI has introduced new threat vectors such as deepfakes, synthetic spams, and more

45%

Gen AI has made harmful content harder to detect

45%

Gen AI has enabled abuse at scale, including bot-generated trolling, frauds, and more

37%

Gen AI has complicated enforcement and surfaced ethical dilemmas

34%

Gen AI has created moderation challenges such as detecting stylized AI-generated avatars and contextually ambiguous interactions

Gen AI's impact is expected to be the highest on UGCs, sandbox environments, immersive virtual worlds, and metaverse platforms where the open-ended nature of content creation and avatar-based interaction exposes platforms to threats such as deepfake avatars, synthetic grooming, and AI-generated harmful assets. Battle royale and tactical shooter and social party games will also be impacted due to real-time voice abuse, impersonation, and offensive content generation.

While moderation is essential for safe online spaces, it should not come at the cost of the gaming experience. For example, sarcasm in intense situations is an inherent human characteristic and should not simply be automatically tagged as offensive content without context. With an expanding array of game titles, gaming enterprises face the conundrum of enhancing security while preserving elements of surprise. The next section outlines strategies gaming enterprises have adopted internally to address these challenges.

Striking the right balance between moderation and authentic gameplay

With the expanding scope and complexity of T&S risks, gaming enterprises are making targeted investments to mitigate harm while preserving the integrity, spontaneity, and authenticity of the gameplay experience. However, the maturity of these efforts ranges widely from ad-hoc, team-specific actions to centralized, fully integrated governance models. According to our survey results, ~32% of the enterprises have mature, structured governance with executive sponsorship and integrated T&S into cross-functional processes, while others remain at early or fragmented stages of development.

Amid this backdrop, enterprises are implementing concrete, organization-wide measures that embed safety into the fabric of their enterprises. Exhibit 6 highlights top actions enterprises have taken to mitigate T&S challenges.

Exhibit 6: Gaming enterprises' leading T&S investments

Source: Everest Group (2025)

79%	Establishing or strengthening centralized T&S governance structures and integrating T&S priorities into cross-functional planning
79%	Developing proactive moderation mechanisms
68%	Embedding Safety by Design early across products in design phase
61%	Improving T&S through PX-related measurable outcomes
53%	Launching player education initiatives on digital safety, reporting tools, and responsible gameplay
29%	Ensuring regulatory and compliance readiness

Establishing or strengthening centralized T&S governance structures and integrating T&S priorities into cross-functional planning

Gaming enterprises are strengthening T&S governance through dedicated structures to ensure centralized oversight and clear accountability. They are implementing robust internal policies, including player guidelines, event-specific codes of conduct, and tiered enforcement frameworks such as three-strike systems. They are also strengthening in-house T&S teams and recruiting specialized gaming talent to lead safety initiatives. These teams now function as strategic partners, working closely with game design, legal, engineering, and community functions to embed T&S priorities across the organization and drive cohesive, cross-functional execution. At the same time, enterprises are trying to preserve the integrity of the player journey by ensuring that enforcement is applied with contextual sensitivity and does not affect gameplay fluidity.

Case in point
Roblox has developed a comprehensive safety governance structure that includes clear community standards, age-appropriate content guidelines, and proactive moderation strategies. It has also partnered with organizations such as the International Age Rating Coalition to rate UGC, ensuring that players have access to safe and appropriate gaming experiences.

Developing proactive moderation mechanisms

Leading gaming enterprises are moving beyond reactive user reports and post-incident enforcement by embedding moderation capabilities earlier in the player journey. They are utilizing AI moderation capabilities across key T&S functions, including child safety, toxicity detection, and age verification, identity misuse, and fake account detection. They are also harnessing gen AI to deepen contextual understanding, automate policy enforcement, monitor player wellness, and optimize moderation workflows at scale, enabling faster, more adaptive responses to emerging threats.

Proactive moderation is enabling more context-aware safety systems that can distinguish between disruptive and competitive behavior. Many platforms now allow players to control exposure to mature language or content, ensuring that safety measures support, rather than dilute, the spontaneity and immersion of gameplay.

Additionally, upholding the importance of human moderation for critical and high-context decisions, gaming enterprises implement role-specific training and upskilling teams in emerging areas such as gen AI misuse, multilingual content moderation, and mental health intervention. Enterprises are also institutionalizing wellness programs, offering psychological support and equipping teams with trauma-informed practices.

Embedding Safety by Design early across products in design phase

Gaming enterprises are embedding Safety by Design principles early in development to minimize abuse vectors, enable real-time intervention, and ensure player protection is integrated seamlessly into core gameplay experiences. Some of the key Safety by Design features that enterprises are implementing include:

- Account verification mechanisms to deter impersonation, underage access, and bot creation from the outset (for example, Epic Games' multi-factor authentication for account security)
- Built-in content moderation filters deployed directly in game engines to preempt harmful content (for example, Minecraft's real-time chat filtering and language moderation tools)
- In-game user reporting tools embedded contextually, enabling players to flag violations with minimal friction (for example, Valorant's one-click player report during or after gameplay)
- In-game blocking and muting features that give players control over interactions and exposure to toxic behavior (for example, Rocket League's block and report options during mid-match)

- Onboarding safety tutorials and UI cues that educate players on community standards, privacy settings, and reporting workflows
- Safety-led design reviews (red-teaming and sandbox testing) as part of product development sprints ensuring T&S risks are evaluated alongside gameplay and engagement (for example, Meta's Horizon Worlds conducting red-teaming exercises for social VR safety)
- Customizable parental controls allowing guardians to limit interactions, playtime, and content types for younger users (for example, Roblox allowing granular content and chat controls for parents)
- Rate-limiting or cooldown mechanisms to reduce spam, harassment, or abuse of key game features (for example, Pokémon Unite imposing trading limits to prevent spam and scams)

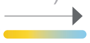
In addition, leading platforms are prioritizing configurability, embedding opt-in filters, flexible language controls, and context-specific muting to allow players to tailor their own safety thresholds, preserving agency and avoiding over-moderation.

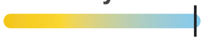
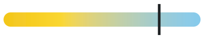
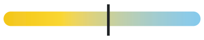
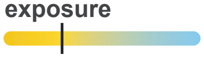
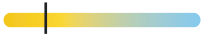
Improving T&S through PX-related measurable outcomes

Gaming enterprises are placing increasing emphasis on PX-linked KPIs to guide decision-making, evaluate safety interventions' effectiveness, and foster long-term player trust. While traditional operational metrics continue to be important, these enterprises are moving beyond traditional models to adopt a more holistic measurement approach, tracking dimensions such as player sentiment and trust, moderation accuracy and consistency, proactivity in intervention, and experience variance across player segments. Exhibit 7 lists the different KPIs tracked and their relevance to gaming enterprises.

Exhibit 7: PX-linked KPIs and relative importance to gaming enterprises

Source: Everest Group (2025)

Enterprise importance Low  High

Operational efficiency 	Volume of safety-related support tickets per 1,000 players	Complaint rate (for example, abuse reports, false positives)	Time to acknowledge player reports (versus full resolution)	Escalation-to-resolution time
Player sentiment and trust 	Regional or demographic variance in safety experience	Player feedback on in-game reporting tools	Moderation net promoter score: trust score, sentiment rating, retention after safety incidents	
Behavior recurrence and deterrence 	Repeat offender rate (recurrence of violations by previously flagged players)			
Harm prevention and exposure 	Toxicity exposure rate (% of players exposed to harmful content)		Volume of proactive safety interventions versus reactive responses	
Accuracy and enforcement quality 	Accuracy rate of moderation and safety tools (false positives/negatives)		Appeal reversal rate (percentage of content/ actions overturned on appeal)	

These PX-linked indicators not only help enterprises detect friction points or gaps in the safety experience but also serve as internal benchmarks for iteration, accountability, and cross-functional alignment.

Launching player education initiatives on digital safety, reporting tools, and responsible gameplay

To enhance safety while preserving PX, gaming enterprises are increasingly investing in player education initiatives that promote digital safety, responsible gameplay, and effective use of platform tools. These initiatives are designed to equip players with the knowledge to recognize risks and take proactive measures without interrupting their in-game journey. Enterprises are embedding in-game tutorials, interactive onboarding modules, and contextual UI prompts that educate players on privacy settings, community standards, and how to report abusive behavior in real time. Simultaneously, enterprises are launching awareness campaigns across community forums, social channels, and events to reinforce positive behavior and explain the rationale behind safety policies. These efforts not only reduce reliance on reactive moderation but also build a culture of shared accountability, where players understand their role in maintaining a respectful and inclusive environment – all while ensuring that the core gameplay remains fluid, intuitive, and engaging.

Ensuring regulatory and compliance readiness

To align with evolving global regulatory standards, gaming enterprises are issuing transparency reports, disclosing platform responses to harmful content, and detailing enforcement metrics such as content takedowns, user reports, moderation actions, and error rates. Privacy policies are being revised to account for AI-driven content review, data usage, and algorithmic enforcement decisions. Enterprises are also building compliance frameworks that cover a broad regulatory landscape, including the EU’s Digital Services Act (DSA), the Children’s Online Privacy Protection Act (COPPA), the General Data Protection Regulation (GDPR), and associated laws around age assurance, financial crime and compliance guidelines, and anti-money laundering in real-money gaming. In addition to building robust consent workflows and back-end controls, leading platforms are increasingly focused on educating players, explaining what these regulations entail, why they matter, and how platform changes are driven by the need to protect users and remain compliant.

Case in point
Miniclip publishes annual transparency reports detailing its management and monitoring of content and conduct policies. It has also updated its privacy policy to address emerging data usage practices related to AI-enabled content moderation and has implemented user-centric safeguards such as parental consent workflows to enhance user trust and ensure compliance with GDPR and COPPA.

While gaming enterprises continue to take T&S challenges head-on, many acknowledge that internal efforts alone are no longer sufficient to address the scale, speed, and evolving complexity of the threat landscape. Even with stronger governance frameworks and more proactive interventions, gaps remain, particularly in moderating high-risk content, ensuring consistent global enforcement, and sustaining resilient, round-the-clock operations. Exhibit 8 highlights the primary challenges enterprises face while scaling T&S internally.

Exhibit 8: Challenges gaming enterprises face while scaling T&S internally

Source: Everest Group (2025)

% of respondents selecting these as top five challenges

68% Challenges in staying compliant with fast-evolving global regulations (e.g., DSA, GDPR, COPPA)

66% Difficulty enforcing safety without disrupting PX

66% Unclear roles and fragmented accountability across T&S, policy, product, and legal functions

58% Operating with limited budgets, leadership support, or team alignment

55% Inconsistencies in enforcing policies fairly and transparently across regions and teams

As in-house-only T&S models come under increasing strain, gaming enterprises are turning to providers to augment capacity, enhance responsiveness, and access specialized expertise.

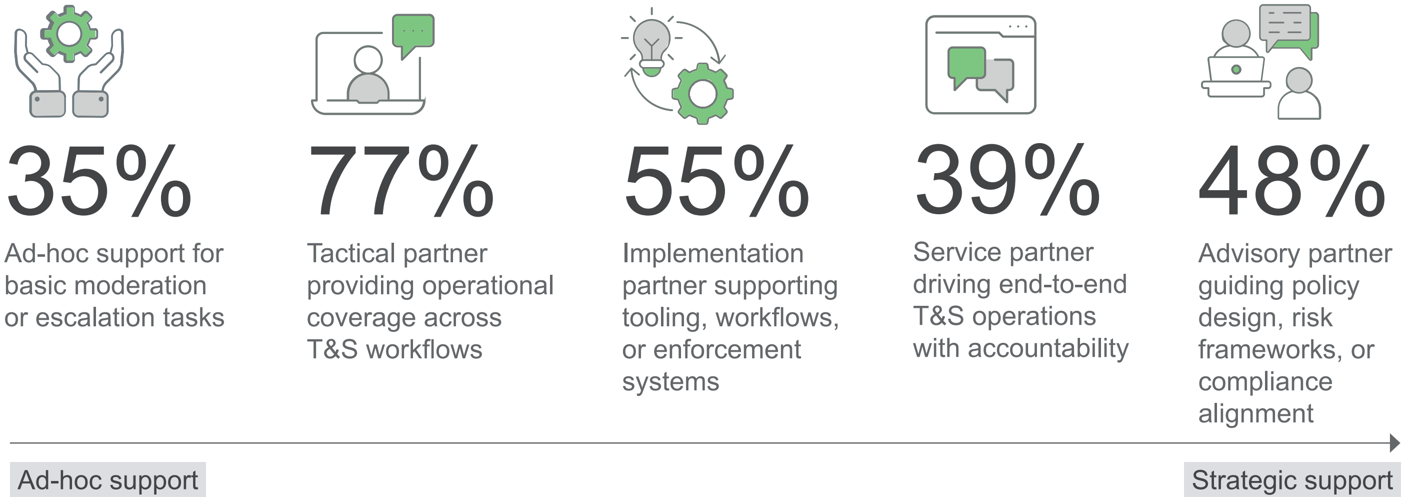
The current T&S outsourcing landscape

Understanding the role of providers

The role of providers varies significantly across enterprises, shaped by the maturity and scale of internal T&S systems, geographic and regulatory complexities, and the nature of platform-specific risks. Accordingly, engagement models span from transactional to strategic, and from modular support to end-to-end managed services, depending on each platform's unique risk profile and operational needs. Exhibit 9 outlines how enterprises perceive the role of providers in their T&S operations.

Exhibit 9: Role of providers in gaming T&S operations

Source: Everest Group (2025)



35% Ad-hoc support for basic moderation or escalation tasks

77% Tactical partner providing operational coverage across T&S workflows

55% Implementation partner supporting tooling, workflows, or enforcement systems

39% Service partner driving end-to-end T&S operations with accountability

48% Advisory partner guiding policy design, risk frameworks, or compliance alignment

Ad-hoc support → Strategic support

Providers support a wide range of T&S functions. Scaled moderation is among the most frequently outsourced areas, particularly for high-volume platforms. Enterprises also rely on partners for incident responses, report triage, appeal handling, and policy enforcement, where both speed and contextual judgment are essential. Exhibit 10 highlights the most outsourced functions in gaming T&S today.

Exhibit 10: List of top outsourced T&S functions in gaming

Source: Everest Group (2025)

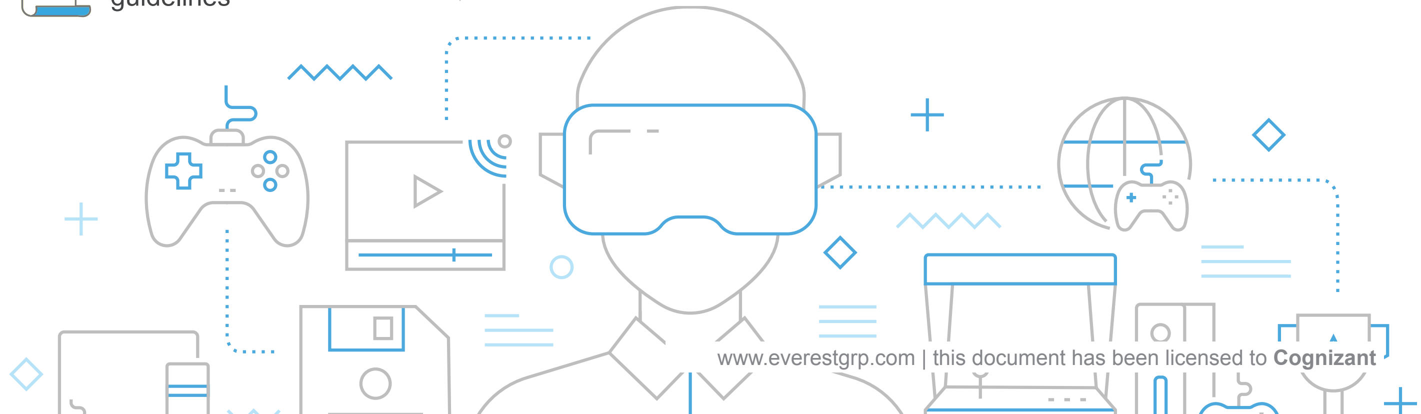
58% **Player report triage** and high-risk escalation, including handling abuse reports, severity classification, and urgent incident escalation

42% **Scaled moderation** across formats (regional or multilingual moderation and cultural context review)

42% **Compliance tracking and documentation** (for example, DSA, GDPR, COPPA)

42% **Applying and enforcing platform policies** based on established T&S guidelines

39% **T&S data labeling and model training**, including annotating harmful content to train AI moderation and detection models

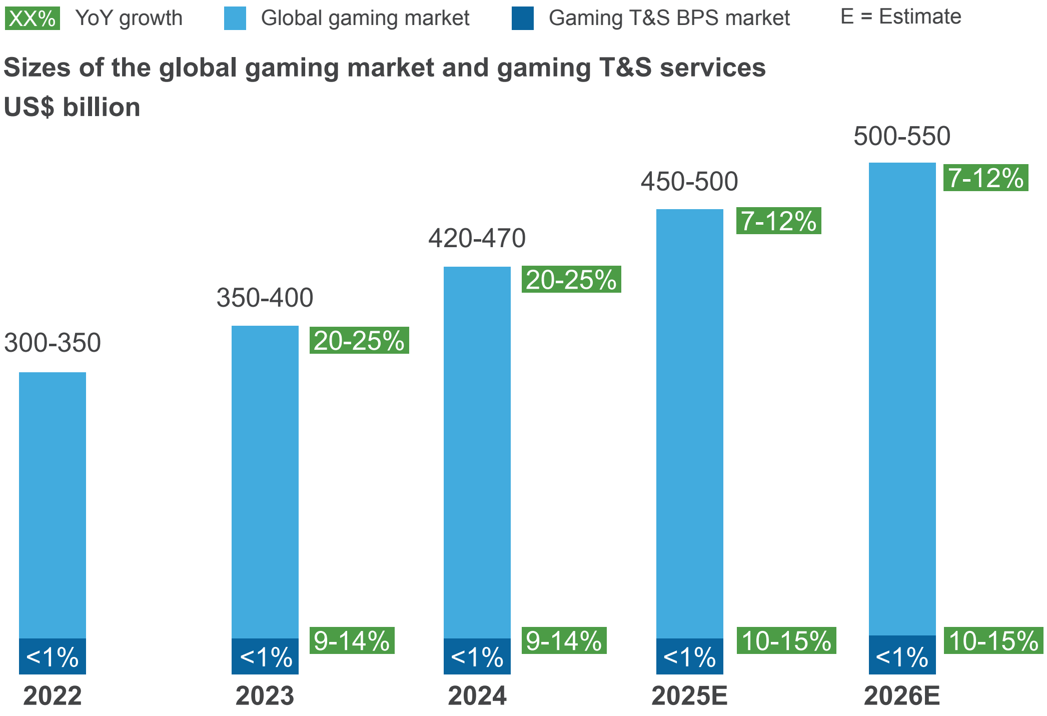


Factors inhibiting outsourcing to providers

While enterprises increasingly recognize the value of outsourcing to scale their T&S operations, many remain cautious about fully committing to external partnerships. Over the past few years, the global gaming market has expanded rapidly, driven by rising user engagement and content creation. However, outsourcing for T&S has not kept pace. Adoption remains limited, with T&S BPS spending still accounting for less than 1% of total industry size. This disparity, as shown in Exhibit 11, highlights a significant gap between potential and actual investments.

Exhibit 11: Global gaming market and gaming T&S services market sizes

Source: Statista and Everest Group (2025)



Despite the advantages of outsourcing, concerns persist around control, transparency, and strategic alignment, shaping how and whether, these partnerships are structured. The following factors make enterprise hesitant to outsource:

- **Loss of control:** concern that external providers lack the cultural understanding and contextual awareness needed to moderate effectively within nuanced gaming environments
- **Transparency and trust issues:** concerns around data privacy, IP protection, and limited visibility into how moderation decisions are made and reviewed
- **Innovation gap:** perception that outsourcing partners may not possess the agility, creativity, or domain-specific expertise required to keep pace with emerging T&S threats

- **Real-time responsiveness concerns:** doubts about whether external teams can deliver the immediacy and precision required to handle live incidents or crisis-level moderation

These reservations are especially pronounced among platforms with distinct community dynamics or highly sensitive content environments. Even among the platforms that outsource, operational challenges frequently arise, impacting the consistency, quality, and scalability of T&S delivery. Exhibit 12 illustrates the most common challenges enterprises encounter when working with providers for T&S operations.

Exhibit 12: Challenges gaming enterprises encounter when working with providers

Source: Everest Group (2025)

1 Integration challenges with internal teams

2 Misalignment on escalation criteria, enforcement policies, or KPIs

3 Lack of sufficient understanding of regional/ cultural context, gaming culture, or abuse patterns

4 Inability to respond quickly to high-severity or time-sensitive incidents

5 Inflexibility in adapting to new or evolving threat types

5* Inadequate support systems to protect moderator well-being and prevent burnout

How outsourcing can be leveraged as a strategic lever to reduce TCO

While concerns around outsourcing are valid, outsourcing remains relevant to reduce TCO while preserving T&S effectiveness. Providers offer proven capabilities in moderating complex formats, deploying scalable AI-human hybrid models, and managing multilingual regional workflows – capabilities often too costly to build in house.

As gaming platforms increasingly face risks that mirror those seen in social media, providers with mature T&S practices in adjacent sectors bring transferable insights and operational depth. They help enhance AI accuracy by feeding real-world moderation data into model refinement pipelines. They also contribute to higher-quality enforcement through structured QA processes, ensure regulatory alignment to reduce the risk of fines or enforcement actions, and support faster adaptation to emerging threat vectors that in-house teams may not be resourced to address alone. Even as enterprise budgets shrink, strategic outsourcing can help reduce TCO while unlocking measurable benefits in efficiency, responsiveness, and safety coverage.

*Rank 5 is shared by two challenges that received an equal number of mentions from respondents

Complementing existing capabilities through third-party support

Enterprises and providers need to collaboratively define and continuously refine their shared responsibilities. By codeveloping governance models, enterprises can retain oversight of strategic decision-making, cross-functional coordination, and organizational context, while providers bring operationally intensive and scale-dependent expertise such as moderation across content formats, appeal handling, data labeling for AI training, and player sentiment analytics. This integrated approach reduces redundancy, ensures consistency in enforcement decisions, and creates a unified safety vision that effectively leverages both internal insights and external innovation.

To execute this integrated approach, enterprises must focus on selecting and structuring partnerships that deliver not just coverage, but impact. Outsourcing expectations must evolve from baseline requirements toward real-time responsiveness, transparent and explainable decision-making frameworks, and alignment with internal policy, product, and compliance teams. Exhibit 13 highlights the top differentiators that enterprises value most when evaluating T&S providers.

Exhibit 13: Top gaming enterprise expectations when evaluating T&S providers

Source: Everest Group (2025)

% of respondents selecting these as top five expectations

81% Deep expertise in gaming-specific safety risks and community norms

61% Expertise in moderating voice, streams, multiplayer chats

58% Scalable, multilingual safety with local nuances

55% Transparent reporting and explainable decision-making frameworks

52% Demonstrated impact on player trust, sentiment, or churn reduction

With these differentiators in mind, Everest Group proposes the DEFEND framework that covers the best practices on how enterprises can leverage providers effectively to complement internal functions, reinforce safety frameworks, and scale responsibly, as highlighted in Exhibit 14.

Exhibit 14: Everest Group's DEFEND framework for enterprise best practices for outsourcing

Source: Everest Group (2025)

D	**Design modular delivery models**	Segment T&S needs by content type, risk intensity, and game format and co-create modular outsourcing models tailored to these profiles, including developing flexible SLAs, tiered response protocols, and service scopes that vary across genres (for example, fast-paced shooters versus sandbox UGC) Ensure providers can scale selectively and adapt quickly to live ops environments and seasonal player spikes
E	**Enable hybrid human + AI enforcement**	Integrate providers into the current AI-human moderation ecosystem, giving providers structured roles in red-teaming, model evaluation, and feedback data loops; for non-gen AI workflows, ensure clarity on queue prioritization, task routing, and manual escalation procedures Internal teams should align with providers on automation thresholds, escalation triggers, reviewer training protocols, and AI explainability standards, especially for real-time content such as live streams
F	**Formalize escalation and appeal workflows**	Develop joint incident response frameworks with providers, including scenario-based escalation playbooks (for example, hate raids, grooming, fraud) Allow provider teams access to tooling and decision frameworks for first-line handling, while internal teams retain high-severity final calls Use policy logs and moderation evidence to jointly audit user appeals and decision reversals
E	**Enforce QA and regulatory alignment**	Define T&S policies centrally and work with providers to localize interpretation and maintain execution fidelity, including setting up QA scorecards, calibration sessions, and pre-audit documentation pipelines Proactively track regulatory developments (DSA, COPPA, GDPR) and ensure provider teams are briefed and equipped for compliant implementation Leverage providers to support policy audits, identify grey areas, benchmark enforcement standards, and stay ahead of evolving regulatory requirements
N	**Nurture integration across safety, product, and PX**	Loop providers for cross-functional collaboration, with providers participating in product roadmap reviews, policy update briefings, and PX sentiment analyses, leveraging them as collaborative partners Internal teams can also partner with providers to identify new threat vectors, simulate abuse scenarios, and track shifts in player behaviors
D	**Deliver insights through real-time reporting and feedback**	While internal teams can drive T&S intelligence, providers should be used to feed data from moderation workflows into shared dashboards, enabling faster decision-making, better policy tuning, and proactive community interventions by internal teams

With these best practices, enterprises can ensure that they simply do not outsource for scale, but to strategically embed providers across the value chain where they can deliver measurable impact.

Future of T&S in gaming

As safety expectations evolve alongside increasingly complex platform dynamics, gaming enterprises are fundamentally rethinking the role of T&S. No longer viewed as a purely operational function, T&S is steadily moving toward becoming a strategic pillar, embedded across product, design, policy, and legal workflows. According to our survey, nearly 34% of gaming enterprises expect their T&S function to evolve into an integrated capability embedded across product, design, policy, and legal, while 16% envision it becoming truly transformational, leading innovation in AI safety, red-teaming, and proactive risk mitigation. This evolution reflects a broader shift in mindset: from reactive enforcement toward proactive, integrated, and innovation-led safety ecosystems.

Looking ahead, the emergence of agentic AI systems capable of making autonomous decisions and adapting based on evolving context will reshape the T&S landscape in gaming. Unlike traditional AI tools that rely on fixed rule sets or supervised models, agentic AI can proactively assess risk, personalize interventions, and optimize enforcement decisions at scale. To prepare, enterprises should build explainability frameworks and establish AI ethics protocols that ensure fairness, auditability, and transparency in automated decision-making. Enterprises will also need to align product, legal, and safety teams to develop new governance models for agentic systems, balancing efficiency with accountability as AI evolves from assistive to autonomous in the T&S function.

Games are evolving and so are the T&S requirements. Technology is changing and provides both an added layer of complexity to T&S operations as well infinite scope of innovation and differentiation. Embracing the evolution proactively and embedding safety as a core principle will drive T&S operations of the future.

Everest Group®
With you on the journey

Everest Group is a leading research firm helping business leaders make confident decisions. We guide clients through today's market challenges and strengthen their strategies by applying contextualized problem-solving to their unique situations. This drives maximized operational and financial performance and transformative experiences. Our deep expertise and tenacious research focused on technology, business processes, and engineering through the lenses of talent, sustainability, and sourcing delivers precise and action-oriented guidance. Find further details and in-depth content at www.everestgrp.com.

This study was funded, in part, by
Cognizant

For more information about
Everest Group, please contact:

+1-214-451-3000
info@everestgrp.com

David Rickard, Partner
david.rickard@everestgrp.com

Abhijnan Dasgupta, Practice Director
abhijnan.dasgupta@everestgrp.com

Dhruv Khosla, Practice Director
dhruv.khosla@everestgrp.com

Tushar Pathela, Senior Analyst
tushar.pathela@everestgrp.com

Nimisha Awasthi, Analyst
nimisha.awasthi@everestgrp.com

Notice and Disclaimers

Important information. Please read this notice carefully and in its entirety. By accessing Everest Group materials, products or services, you agree to Everest Group's Terms of Use.

Everest Group's Terms of Use, available at www.everestgrp.com/terms-of-use, is hereby incorporated by reference as if fully reproduced herein. Parts of the Terms of Use are shown below for convenience only. Please refer to the link above for the full and official version of the Terms of Use.

Everest Group is not registered as an investment adviser or research analyst with the U.S. Securities and Exchange Commission, the Financial Industry Regulation Authority (FINRA), or any state or foreign (non-U.S.) securities regulatory authority. For the avoidance of doubt, Everest Group is not providing any advice concerning securities as defined by the law or any regulatory entity or an analysis of equity securities as defined by the law or any regulatory entity. All properties, assets, materials, products and/or services (including in relation to gen AI) of Everest Group are provided or made available for access on the basis such is for informational purposes only and provided "AS IS" without any warranty of any kind, whether express, implied, or otherwise, including warranties of completeness, accuracy, reliability, noninfringement, adequacy, merchantability or fitness for a particular purpose. All implied warranties are disclaimed to the extent permitted by law. You understand and expressly agree that you assume the entire risk as to your use and any reliance upon such.

Everest Group is not a legal, tax, financial, or investment adviser, and nothing provided by Everest Group is legal, tax, financial, or investment advice. Nothing Everest Group provides is an offer to sell or a solicitation of an offer to purchase any securities or instruments from any entity. Nothing from Everest Group may be used or relied upon in evaluating the merits of any investment. Do not base any investment decisions, in whole or part, on anything provided by Everest Group.

Everest Group materials, products and/or services represent research opinions or viewpoints, not representations or statements of fact. Accessing, using, or receiving a grant of access to Everest Group materials, products and/or services does not constitute any recommendation by Everest Group to (1) take any action or refrain from taking any action or (2) enter into a particular transaction. Nothing from Everest Group will be relied upon or interpreted as a promise or representation as to past, present, or future performance of a business or a market. The information contained in any Everest Group material, product and/or service is as of the date prepared and Everest Group has no duty or obligation to update or revise the information or documentation.

Everest Group collects data and information from sources it, in its sole discretion, considers reliable. Everest Group may have obtained data or information that appears in its materials, products and/or services from the parties mentioned therein, public sources, or third-party sources, including data and information related to financials, estimates, and/or forecasts. Everest Group is not a certified public accounting firm or an accredited auditor and has not audited financials. Everest Group assumes no responsibility for independently verifying such information.

Companies mentioned in Everest Group materials, products and/or services may be customers of Everest Group or have interacted with Everest Group in some other way, including, without limitation, participating in Everest Group research activities.

A global study on gaming trust and safety by Everest Group, supported by Cognizant