WebPurify
An Intouch CX Company

WEBPURIFY E-BOOK

From friction to flow:
Solving common CX pain points with Trust & Safety

How moderation and smart collaboration between teams can transform customer experience

EXECUTIVE SUMMARY

Even the best customer experience can fall apart in an instant – if a user is scammed, harassed, or loses trust in your platform. What many companies overlook is that these are not just safety failures – they are customer experience failures. When trust breaks down, so does retention.

As platforms grow more complex and user expectations rise, the line between protecting users and delighting them has all but disappeared. A scammer slipping through moderation, a flood of fake reviews, or a mishandled abuse report is more than just a safety failure; it's a customer experience failure.

Our latest ebook explores how integrating trust and safety into the heart of your CX strategy can turn moments of friction into opportunities to build loyalty. Drawing on insights from three leading experts who joined WebPurify's – an IntouchCX company – webinar, *Turning Trust & Safety into a CX Advantage*, you'll hear from:

Francis Stones
Global Head of Brand Safety, TikTok

Deniz Alkan
Director of Reviews, Trustpilot

AJ Brown
Chief Operating Officer, Brand Safety Institute

Together, they offer practical and strategic guidance on how brands can use trust and safety not just to mitigate harm, but to provide faster support, drive loyalty, protect brand equity, and deliver seamless, "invisible" user journeys. Along with WebPurify's Head of Trust & Safety for EMEA, Ailís Daly, they'll explore:

- The most common **CX friction points** caused by poor moderation, including scams, abuse, and slow support
- The hidden costs of unmoderated platforms, from PR fallout to agent burnout

- How automation and AI (when paired with human oversight) can improve both safety and efficiency
- Real-world examples of moderation systems that reduced churn, improved CSAT, and cut operational costs
- A framework for breaking down organizational silos between CX, T&S, Legal, Product, and Marketing
- Tactical advice for aligning **metrics, messaging, and escalation paths** to build safer, more trusted platforms

The key takeaway here is very simple: trust and safety isn't a back-office function. It's a strategic lever for growth, loyalty, and customer retention. When every team shares responsibility for building trust, user experience flows.

WHAT IF TRUST & SAFETY WAS YOUR BEST CX HIRE?

Even when people love a company, **17% will walk away after just one bad experience.** And if that bad experience is encountering a scammer, a flood of fake reviews, or harassment on your platform, you can bet those customers won't be back. Customer experience (CX) and Trust & Safety (T&S) are often treated as separate silos – one team delighting customers, the other policing content and behavior – but the reality is that a glitch in safety is a CX failure.

But with risk comes opportunity, for while **T&S issues = CX issues**, solving them can prove to be your biggest differentiator in customer experience.

What if, instead of seeing T&S as a compliance checkbox, you treated it as your next star CX hire? This ebook builds on insights shared in our recent webinar, *Turning Trust & Safety into a CX Advantage*, hosted by Ailís Daly, *Head of Trust & Safety for EMEA at WebPurify (an IntouchCX company)* and featuring Francis Stones, *Global Head of Brand Safety at TikTok;* Deniz Alkan, *Director of Reviews at Trustpilot;* and AJ Brown, *Chief Operating Officer at the Brand Safety Institute.*

We'll examine:

- Common CX pain points – such as fraud, fake accounts, toxic content, moderation bottlenecks – and show how integrating strong, consistent T&S practices can smooth them out.

- How forward-thinking brands turn friction into flow by designing safer products upstream, not scrambling to fix issues downstream.

- We'll share real examples (like fraud detection that cut handling time by **71%**, or AI assistance that boosted first-contact resolution and CSAT), and hear from industry leaders on bridging the gap between T&S, CX, Product, Legal, and Marketing.

By the end of this guide, you should walk away with actionable strategies to break down silos in your business and start delivering a seamless, "invisible UX" where users feel safe and supported without even noticing the behind-the-scenes work.

"Trust and safety is not something that should be a second thought or something that comes after you actually get the customer – it should be part of your core business strategy."

DENIZ ALKAN, HEAD OF REVIEWS AT TRUSTPILOT

In other words, safety by design. *"If Trustpilot didn't have a good T&S team and the principles that we do, it would be filled with fake reviews and that would make the platform obsolete. We want people to trust the reviews they read and the experiences people share."*

The friction points undermining CX

Every CX leader knows the pain points that frustrate customers and erode loyalty. What they may not realize is how often these issues stem from T&S failures. From fake accounts to slow moderation, even seemingly isolated issues can have knock-on effects that damage CX. Below, we explore the most common friction points and how they trace back to T&S weaknesses.

WHAT USERS SEE:
Fake accounts and bad actors

Globally, fake online reviews alone are estimated to be a **$152 billion** problem. A staggering **85%** of consumers suspect that reviews they read are "sometimes or often fake," and more than a quarter say just one fake review would make them distrust a company entirely.

Whether it's bot profiles inflating follower counts, cloned accounts mimicking users, or fake reviews propping up scam sellers, this is damaging – but particularly for review platforms, dating apps and marketplaces, where credibility is essential. The moment users feel that other users aren't real, or that feedback is manipulated, engagement drops.

85%
of consumers suspect that reviews they read are "sometime or often fake"

WHAT USERS FEEL:
Abuse, harassment, and harmful content

Toxic content, whether it's hate speech, harassment, bullying, or misinformation, has an immediate emotional impact on users. It makes them feel unsafe, unwelcome, and more likely to disengage.

As Deniz says: *"You're going to lose users and no one's going to want to come to your platform and engage with your content if they're being trolled online or they're reading stuff that they don't see as the truth. I don't see how that's not a customer retention issue."* Allowing harm to continue unchecked is, effectively, a churn strategy.

WHAT USERS EXPERIENCE:
Scams and support breakdowns

When a T&S failure occurs – for instance, a fraudulent seller on a marketplace or a phishing scheme via fake accounts – users become victims. Their accounts are compromised, money is lost. But the fallout doesn't end there.

Support teams often bear the brunt of these failures. One scam can trigger hundreds or thousands of support tickets, overwhelming agents and increasing handle times. Customers dealing with fraud or safety issues often face slow resolution, unclear communication, and inconsistent outcomes. The result is a compounding of negative experiences that damages long-term loyalty.

When trust breaks down in transactional environments, users quickly turn to support for resolution. Fake listings, phishing messages, and payment fraud can lead to a flood of complaints. This increases handling times and stretches CX teams thin. Worse still, victims often blame the platform for the incident, even if technically the bad actor was a third-party user.

HARMFUL CONTENT ERODES BRAND SAFETY AND PUBLIC TRUST

The presence of hate speech, misinformation or extremist content will also damage your public perception. It makes headlines, prompts advertiser boycotts, and drives cautious users away. A platform that doesn't moderate proactively is perceived as irresponsible, even complicit. The damage can be swift and long-lasting.

As Deniz observes, users are increasingly tuned into these risks. *"They're more aware than ever of the role trust and safety plays in their experience. If they feel unprotected, they'll leave your platform."*

THE SUBTLER PAIN POINTS

Not all T&S friction points are obvious, but they still degrade CX over time.

- **Moderation delays:** when violative content sits visible for too long, users lose faith in the platform's ability to protect them.
- **Inconsistent enforcement:** conflicting or unclear policy decisions frustrate users who feel they've been treated unfairly.
- **Lack of safety features:** the absence of in-app reporting tools, for instance, or clear routes to help, leaves users feeling helpless when something goes wrong.

The hidden cost of unmoderated platforms

The effects of poor moderation ripple through every part of a digital business. From bloated support costs to brand damage that drives users away, T&S failures are some of the most expensive mistakes a company can make. And yet, they're often underfunded or deprioritized until crisis hits.

Skyrocketing support volume and resolution times

If you're not preventing fraud, abuse, and other safety issues upstream, they all flow downstream to your support channels. Every fake listing that scams a customer, every hate speech incident that drives complaints – it all ends up as tickets, calls, and chats for your customer service team.

This means higher contact volume, longer average handle times, and a need for more agents. The average cost per support contact can vary across industries, but whatever sector you're in, thousands of extra contacts quickly add up to serious $$$. And that figure doesn't include the opportunity cost of your team being stuck in reactive mode instead of focusing on proactive customer care or upselling.

Unmoderated platforms essentially tax your support operation. (Plus, as we'll discuss, it takes a toll on agent morale to constantly fight fires that could have been prevented.)

Negative brand perception and PR crises

In our digital age, one explosive incident can quickly ignite on social media or even spread to the news front pages. Think of headlines about platforms failing to remove dangerous content, or app stores banning apps over lax safety controls. The direct fallout is bad enough (boycotts, protests, regulatory scrutiny), but the indirect hit to customer trust is worse. Users start wondering, "Is this platform safe for me? Do they care about protecting us?"

As Deniz warns, T&S shouldn't be seen as a blocker to growth, but rather insurance against catastrophe: *"If you launch a feature that is misused in the worst possible way and you end up on the front page of a newspaper, then you lose many more customers or users than you would have if you were just taking your time and going through proper T&S reviews."*

In other words, an ounce of prevention is worth a million dollars of cure (or in some cases, billions – just ask platforms that faced advertiser boycotts over unsafe content). Brand trust, once lost, is extremely hard to rebuild.

Costly compliance risks

Regulatory pressure is mounting in the wake of legislation such as the EU's Digital Safety Act. New laws around child safety, data privacy, and harmful content in general mean that failures to act can result in fines or court battles, not just headlines. Compliance is no longer optional.

Companies that approach T&S as a proactive discipline are better positioned to meet evolving legal standards. And in many cases, the cost of compliance is far lower than the price of remediation after something goes wrong.

Emotional toll on employees and community

There's also a human cost to consider. Behind every flagged post is a human who had to look at it.

When your frontline staff and moderators are constantly dealing with graphic violence, child exploitation content, threats, or distraught victims of fraud, it wears them down.

Even support agents, while usually not reviewing graphic media, may have to help scam victims or handle angry, abusive user reports. The burnout is very real and ultimately leads to higher turnover and lower quality service. And if your platform leaves users to fend for themselves, the community as a whole can become toxic, driving away the very customers you want to keep.

Brand partnerships at stake

Advertisers are also increasingly risk-averse. No one wants their campaign appearing next to hate speech or misinformation. A misaligned ad placement may cost a large brand a few thousand in wasted impressions – a storm it can probably weather. But for a small business, this type of loss could be catastrophic.

"You can't have growth without good trust and safety," says Francis Stones, Head of Brand Safety and Industry Relations at TikTok. *"Our advertisers know this. They want to be part of safe environments because brand equity is on the line."*

Smart brands increasingly view safety as part of the value exchange, and platforms that fail to offer safe, curated environments risk losing high-value partnerships.

"You can't have growth without good trust and safety. Our advertisers know this. They want to be part of safe environments because brand equity is on the line."

FRANCIS STONES, *HEAD OF BRAND SAFETY AND INDUSTRY RELATIONS AT TIKTOK*

Efficiency gains from smarter systems

Done right, T&S can make everything else work better. In one IntouchCX case study, a beauty brand using AI-enabled agent assist saw first contact resolution improve by **17%** and CSAT climb **10%**.

Deniz notes that by automating low-level triage and outsourcing where appropriate, companies can reduce T&S costs while improving scale and accuracy. *"That initial investment can pay off very quickly.*

"Investing in T&S doesn't necessarily mean the amount of money you put into T&S teams is going to go up. It might be that you invest more at the beginning and then everything else is optimized because of that initial investment."

"We've conditioned people to think of trust and safety just as policy enforcement, the team that shows up when something goes wrong," says AJ Brown, Chief Operating Officer, Brand Safety Institute. *"But now, it's equally important to bake T&S work into upstream measures, designing products, systems, and policies that prevent harm from occurring in the first place."* Ultimately, the hidden cost of poor moderation is not what you spend, but what you lose. Customers. Credibility. Control.

T&S as a CX power tool

Imagine if your best customer service agent worked 24/7, behind the scenes, to silently resolve issues before they ever reached your users. That, in essence, is what a strong T&S function can do.

When T&S works well, users barely notice it. There are no red flags, no abusive messages, no suspicious transactions or policy violations. Everything simply works. This is the ideal of what some call "invisible UX": the behind-the-scenes systems that keep users safe without ever interrupting the journey.

But this kind of seamlessness isn't accidental. It's carefully built! And increasingly, it's powered by proactive moderation, automation, and AI, which work together to pre-empt countless support issues and ensure customers have smoother experiences.

Let's look at a few ways T&S tactics directly translate into smoother CX:

Pre-empting support issues through proactive moderation

When moderation is reactive, CX agents are left cleaning up messes, such as fielding complaints, responding to abuse reports, and trying to restore broken trust.

But with the right T&S tools in place, many of those issues never reach the user.

"You could have an AI go through hundreds of thousands of comments and give you the top five issues that people are talking about, or you could have an AI tailor a more custom response to users, so you get that balance of scale versus making sure that a customer is feeling heard," Deniz says.

Automated content moderation and filtering

Modern AI and machine learning tools can detect and remove the vast majority of policy-violating content (hate speech, nudity, threats, spam) before anyone even sees it. For example, **YouTube reports that more than 80% of the videos they remove for violating policies are first flagged by automated systems** – and many are removed before getting a single view.

YouTube reports that more than 80% of the videos they remove for violating policies are first flagged by automated systems

The same goes for text, images, and chats: automated filters and human moderators working in tandem, a model which WebPurify has pioneered, can catch bad content so fast that your average user rarely (if ever) encounters the worst stuff, and doesn't have to personally report it. They just enjoy the platform without incident.

One dating app that implemented AI-based image moderation saw abusive content complaints drop by **60%**, as offensive images were flagged or blurred automatically instead of reaching users' inboxes (case study later in this chapter). By filtering out toxic content and fake posts up front, you protect customers and reduce the burden on support.

But Deniz also stresses the importance of human oversight: *"Having a human in the loop at some stage wherever you use AI is key because trends change over time and you can't just train a model and expect it to perform the same way without degrading. It's really important to keep that training data fresh, and keep humans in the loop for any issues that might pop up or any trends that might be missed by AI."*

Automation gets you scale. Human review ensures context, empathy, and adaptability.

Fraud detection and prevention systems

Similarly, advanced fraud detection can stop scammers and bad transactions in their tracks. Machine learning models can identify patterns (multiple high-value orders, suspicious IPs, etc) far faster than manual reviews.

At WebPurify, fraud and risk initiatives are now led by Elena Michaeli, a seasoned expert with over 15 years' experience in fraud prevention, compliance, and cybersecurity, including time with the Israeli Defence Forces' elite Unit 8200 and major banking institutions. Her background includes building a fraud "fusion center" – integrating fraud, cybersecurity, and compliance into a unified strategy – and she brings that same holistic approach to WebPurify's clients.

"Fraud prevention can't live in a vacuum. It has to work hand-in-hand with cybersecurity, engineering, moderation, customer care – all of it. That cross-functional collaboration is the only way to detect patterns early, adapt fast, and respond with the right actions," Elena says. This philosophy underpins the work being done across WebPurify and IntouchCX, where T&S teams are designing smarter, more agile fraud systems.

One example: **IntouchCX recently ran a pilot with an AI-driven fraud check system for a delivery service, and the impact on efficiency was dramatic:** they **cut average handle time (AHT) by 43% within two months**, and ultimately achieved a **71% reduction in AHT** by the end of the pilot. Customers got their orders verified and issues resolved in a fraction of the time, simply because AI handled the heavy lifting of fraud review.

In another program, **IntouchCX's automation reduced net handle time by 45% while also cutting 30–40% of the cost of fraud operations.** Faster fraud checks = quicker approvals or flags, less waiting for customers, and far fewer cases escalated to specialist teams.

When fraud and security processes are streamlined, legitimate customers face less friction (fewer false declines, shorter hold times) and fraudsters are weeded out early, sparing everyone a lot of grief.

71%
reduction in average handle time

30–40%
reduction in cost of fraud operations

AI-assisted customer support

While we often think of T&S as removing bad actors, it also means empowering the good actors (support agents and community managers) to better help customers. AI can be a co-pilot for agents, providing real-time suggestions, knowledge search, and even automated responses for routine issues. This speeds up response times and ensures consistent, accurate information, which is critical in sensitive situations (such as safety incidents or account security problems).

For instance, **IntouchCX deployed an AI agent-assist tool** (Sidd Pro) with a leading beauty brand's support team. The results were telling: First Contact Resolution **improved by 17%**, agent productivity went up nearly **50%**, and customer satisfaction (CSAT) scores rose by **10%**.

17%
**increase in
First Contact Resolution**
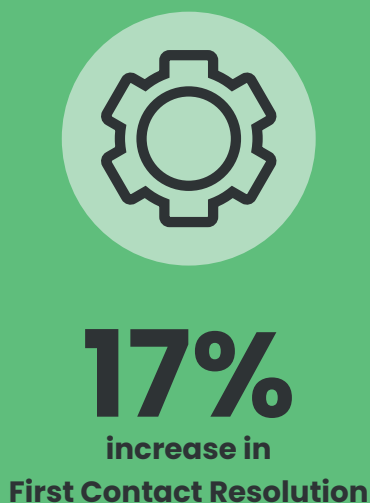
By arming agents with smarter tools, more customers got their issues resolved on the first try, and walked away happier. From a trust perspective, it also means that when a user reaches out about any issue, they get a fast and effective answer – reinforcing their confidence in the platform.

But AJ adds that brands should also be sensitive to public perceptions about the growing role AI is playing in society.

"Consumers are still very skeptical of AI and they're increasingly enraged when companies replace humans with AI," he says. *"Companies shouldn't take for granted that their customers want AI for AI's sake. The improvement needs to be really compelling from a customer experience perspective when you introduce AI, and if it's seen as something that's done because it's trendy or because it'll save money, companies will see and will deserve [the backlash]."*

Invisible UX: safety as a product feature

The examples above show how T&S can serve as a power tool for CX by reducing abusive content, catching fraud early, and speeding up support. All of it removes friction from the user journey. Perhaps the best outcome is when it's so effective the user doesn't even realize it's there. This is the idea behind the concept of "invisible UX."

When a platform "just works," users feel safe and satisfied but don't consciously think about why. They might not notice the AI moderating their feed or the fraud filter shielding them – they just experience a smooth, trustworthy service. Achieving this invisible UX requires diligent, upstream effort from T&S teams, but the payoff is huge.

How do you create an invisible (but rock-solid) safety net? It boils down to clarity, consistency, and the speed in which you enforce policies behind the scenes. If someone breaks the rules or a user flags an issue, swift and consistent action is what builds trust.

In other words, you prove to users that you've got their back by doing the right thing quickly and reliably every time, not just by making empty promises in a policy page or sending out a PR statement. For example, if you report a harassing comment and see it removed and the offender handled within minutes, you feel validated and safe. If you report it and nothing happens for days (or at all), you learn that the platform doesn't really enforce its own rules. It's a huge trust killer.

Consistency also means no favorites or double standards. Apply your rules fairly, whether the offender is a new user or your top influencer. Another aspect of invisible UX is building safety into the product design itself. Consider platforms built around peer-to-peer interactions – like home sharing or ridesharing – where the entire model hinges on trust between strangers. These businesses have had to bake safety and support into the user experience from day one, using product design as their frontline of trust and risk prevention.

"What some platforms in the travel and hospitality space do brilliantly is ensure the user never feels like they're being passed around between teams or told 'that's not our department,'" says Ailís Daly, WebPurify's Head of Trust & Safety for EMEA. *"If a guest feels unsafe or a host has a problem, they know exactly where to go: in-app safety centers, emergency contact buttons, a 24/7 hotline, responsive social media support – you name it. The signposts for getting help are clear and well-placed, especially during moments of high stress. It's very straightforward for someone to engage with the platform when something goes wrong."*

Ailís believes this is the gold standard, and it's achievable whether you're a marketplace serving buyers and sellers or a social platform supporting advertisers and creators. "*This success is achieved by internal orchestration that feels invisible but powerful to the customer,"* she says.

The upshot is that even when incidents happen (and they inevitably do), users feel taken care of. The trust isn't in the absence of problems, Ailís says, but in the presence of a reliable safety net. That's invisible UX!

Most users will never see the emergency hotline or the refund policy for scams until they need it, but knowing (even subconsciously) that those measures exist makes them comfortable using the service. When T&S work is woven into your product, it reduces anxiety for your users. They can navigate confidently, which is the ultimate smooth experience.

Breaking down silos: collaboration between CX, T&S, Product, Legal, and Marketing

One of the biggest blockers to turning T&S into a CX advantage (as polled in our recent webinar) is organizational silos. Each team has its own priorities and metrics:

- Product teams decide what gets shipped.
- CX teams handle the fallout when something goes wrong.
- Legal keeps things compliant.
- Marketing shapes the brand voice.

When these departments operate in isolation, or, worse, at odds, inconsistencies show up where they hurt most: the customer journey. A scammer gets greenlit by Product because T&S didn't review that new feature, and CX is left cleaning up. Or Marketing runs a campaign that inadvertently invites a flood of toxic behavior that the T&S team wasn't warned about. To avoid such debacles, companies need to build tighter collaboration and shared ownership of safety and experience.

According to Deniz, *"T&S is sometimes almost seen as the bad guy, preventing innovation or user growth because we want to check new features before launch and scope them out for safety risks. So there's always this push and pull between T&S and product teams."*

Every interaction is a CX interaction. Yet too often, teams operate in isolation, focused only on their domain. This fragmentation leads to disjointed communication, inconsistent enforcement, and frustrated users. When no one owns the full journey, trust erodes at the seams.

To move from friction to flow, companies need to embed T&S into every stage of the product and user lifecycle. Here's how:

Embedding safety into product DNA

Features that aren't scoped for safety risk backfiring spectacularly, alienating users, creating moderation headaches, or even triggering reputational damage. Product and T&S can sometimes seem at odds, with one pushing for speed and the other for caution. But this tension, when well-managed, can be productive.

This is the case for "shifting left:" building safety into the design and development process rather than bolting it on later. Embedding T&S specialists directly into product squads ensures risks are flagged and mitigated before a feature ships. Others route all new features through a structured trust review, so abuse vectors are addressed early without blocking progress.

The mindset shift is key: it's not T&S versus Product, it's Product with T&S. When safety is treated as a core UX feature, trust becomes part of what makes a product work, not what slows it down. A safety-first approach functions as a quality assurance loop, helping teams ship features that meet user needs and uphold brand integrity.

Build a shared vocabulary

Language is often the first barrier to collaboration. T&S teams speak in terms of risk, enforcement, and policy. CX teams talk empathy, speed, and resolution. Marketing speaks in terms of tone, brand voice, and user perception. Without shared language, alignment is impossible.

"One of the biggest... gaps I observed during my time at a platform was in the language used by different teams," AJ says. *"Trust and safety speaks in terms of risk and harm reduction... But when you're dealing with people who are primarily talking to people outside the company, that's not the case."*

Deniz recommends bringing teams together early to align not just on rules, but on how those rules are communicated: *"You need to decide what the voice of your company is going to look like externally. People don't see it as, 'I'm contacting the T&S department.' They just see the brand. That tone of voice should be agreed on from the start."*

A shared vocabulary doesn't mean templated responses. It means clarity about what kind of voice your company wants to project, whether it's helpful and friendly, firm and fast, or something else entirely. Tactical decisions, such as whether to use personalized emails or templated messages, should be made with cross-functional input, balancing brand tone with operational scale. Even transparency reports, Deniz argues, can be reimagined as stories: *"Make it digestible. Short videos, snippets that show what amazing things your team has done."*

Marketing can play a critical role here. They're uniquely positioned to turn T&S work into brand storytelling – through transparency reports, short videos, infographics, or social content ("We blocked 5,000 scam accounts last quarter – here's how."). When users understand the company's safety efforts, it deepens trust and enhances reputation.

"Trust and safety speaks in terms of risk and harm reduction... But when you're dealing with people who are primarily talking to people outside the company, that's not the case."

AJ BROWN, CHIEF OPERATING OFFICER AT THE BRAND SAFETY INSTITUTE

Bringing Legal in early

Legal teams understand regulatory requirements and liability risks – but only if they're looped in before a crisis. When involved upstream, Legal can help shape policies, ensure compliance, and assist with user-facing communication during incidents.

They also lend credibility. Customers might not see Legal's work, but they feel it: in clear user agreements, swift action on illegal content, and a general sense that the company is operating by the book. Legal can also help shape transparency reports in a way that informs users without creating unnecessary legal exposure.

Legal's proactive involvement turns T&S from a compliance box into a business enabler.

Create consistent escalation paths

When issues arise that straddle teams – say, a user appeals a moderation decision, or a harmful post goes viral – companies need clear paths for escalation. Who owns the resolution? Who drafts the public response? How fast can those decisions be made?

Without defined handoff processes, CX teams risk burning out on issues they're not equipped to solve, and T&S teams may not hear about emerging problems until it's too late.

Embedding liaisons, such as a rotating trust team member sitting in on support calls, or a CX lead attending T&S policy meetings, can foment faster knowledge-sharing. For day-to-day coordination, some companies create feedback loops by categorizing user complaints: is it a policy issue, a tone issue, or a broken process? For high-severity incidents, shared "war room" channels across Legal, PR, and CX can prevent delays or conflicting messages.

Another effective approach is to create shared dashboards or routing systems that bring together signals from community feedback, ticket sentiment, and flagged content. These tools help break the bubble between teams, creating a clearer shared view of what matters most.

More mature teams go further: embedding T&S specialists in CX teams to listen to support calls, and inviting support agents into T&S policy planning. These crossovers surface blind spots and reduce cold handoffs. Everyone sees the user journey in full.

Align KPIs across departments

Metrics drive behavior. If CX teams are measured solely on CSAT, and T&S teams on enforcement speed or volume, their goals may clash. Worse, actions that improve safety might appear to hurt CX if the metrics aren't framed correctly.

"We [at TikTok] gather a huge amount of data and conduct extensive research on customer satisfaction, including trust and safety considerations," Francis says. *"We aim to improve these year on year, but ultimately, everything feeds into our global OKRs, KPIs, and performance metrics. It's all aligned under a shared banner – the umbrella of trust."*

Deniz stresses that customer satisfaction shouldn't be seen as separate from T&S.

"Even if you follow every process correctly, people may still be unhappy. But you can still learn from that. Divide up the feedback. Is this a policy issue? A communication gap? A branding problem?"

Shared metrics such as resolution rate, AHT, or even a "safety satisfaction score" (CSAT for users affected by trust-related issues) can align incentives. Some companies are going further, creating a cross-functional "trust index," which is a company-wide KPI that tracks user sentiment around safety and transparency.

These metrics help shift the mindset from "not my department" to "shared responsibility."

Transparency is a brand asset

Safety work shouldn't only show up when something goes wrong. Proactively telling the story of your T&S efforts – through blogs, newsletters, or social posts – can boost confidence and internal morale.

"Full transparency doesn't mean publishing every policy nuance in real time," AJ says. *"Policy decisions, kind of by definition, create winners and losers, and not everyone's going to be satisfied with their interaction with you even if you've made that as frictionless as humanly possible."*

"Giving some transparent information to your users and customers about how much bad content you've prevented is useful to build that trust," Deniz adds. *"User-facing policy needs to be clear, understandable, not using so much legalese so users are aware what the rules are and they're not overly punished for rules that they had no idea that they needed to adhere to."*

One company might publish quarterly transparency reports. Another might post a LinkedIn update: "Our safety team removed 10,000 pieces of harmful content last month." These efforts show users that safety is a priority. And they give internal teams the recognition they deserve.

"Our customers are the brands that advertise on TikTok. They're human first or even parents first, and marketers second," Francis says.

"One of the things I've found quite effective is grassroots marketing. If you do grassroots campaigns, it almost has more impact." Transparency builds resilience. When things go wrong – as they inevitably do – users who understand your safety posture are more likely to trust your response.

Transparency is an opportunity to make the invisible work of T&S visible.

"Giving some transparent information to your users and customers about how much bad content you've prevented is useful to build that trust."
DENIZ ALKAN, HEAD OF REVIEWS AT TRUSTPILOT

Collaboration is cultural, not just procedural

Ultimately, breaking silos is about mindset. Everyone, from the marketing intern to the legal counsel, influences user trust. As Deniz explains, *"T&S and marketing teams are usually not talking to each other. But they should be. Safety is part of your brand."* When T&S becomes everyone's job, CX stops being a department and becomes a shared commitment across all teams.

Conclusion

The key takeaway here is that you don't have to do it all alone. There are specialized partners and services that can help implement these kinds of solutions.

For example, as part of IntouchCX, WebPurify delivers an end-to-end approach to trust and safety, combining AI-powered tools, human moderation, and operational support to help platforms manage user-generated content, reduce fraud, and protect brand integrity. From real-time content filtering to advanced solutions like the Sidd Pro agent-assist tool, we work as one integrated team to support your customer experience goals at scale.

The key, however, is to approach these partnerships not as "outsourcing a problem" but as extending your CX strategy. The most successful companies maintain close collaboration between their internal teams and external vendors, ensuring that T&S goals are aligned with overall CX goals.

For today's online businesses, trust is the currency of CX. You can spend millions on a beautiful interface, clever marketing, and speedy delivery, but if users don't feel safe and respected, they won't stick around.

Conversely, a platform that users trust – where they consistently have positive, safe interactions – will enjoy loyalty that no coupon or gimmick can buy. T&S work is often hard and thankless, but it is foundational to long-term CX success and brand integrity.

We've explored how common CX pain points such as fraud, abuse, and content issues can be remedied (or better yet, prevented) by robust T&S measures. By taking a proactive, collaborative approach – treating T&S as a partner to Product, CX, Legal, and Marketing – companies can transform these friction points into moments of flow.

A fraud caught = a support call avoided. A troll banned = a community member retained. A clear policy enforced = a PR crisis averted. These "invisible wins" accumulate into a superior customer journey.

Finally, perhaps the most important shift is cultural: making T&S part of the core ethos of your company. When leadership champions safety as a value ("We put user safety first, even if it means slowing down a bit"), it empowers teams to collaborate and prioritize accordingly.

And when employees know that preventing harm is creating a better CX, they'll see the bigger picture in their daily work. This mindset turns what used to be friction ("ugh, do we really need a T&S review?") into flow ("yes, let's get T&S input to make this launch great"). The road to this ideal state is a journey, not a flip of a switch. To help you get started, we've compiled a practical checklist for CX leaders looking to better integrate with their T&S counterparts. Use these as thought starters for your own organization.

CHECKLIST:
Bridging CX and T&S – first steps for leaders

Get to know your T&S team: Schedule a meet-and-greet or shadowing session. Understand their main challenges, processes, and metrics. Share yours too. Building personal rapport and empathy is the foundation for breaking silos.

Map the customer journey for trust gaps

Sit down together and go through the end-to-end user journey. Identify where a user might encounter a trust or safety issue (e.g. account creation, content posting, transaction, support contact). Brainstorm preventative measures for each stage. This "red teaming" exercise will highlight where CX and T&S need to collaborate (and often yields quick wins).

Establish a joint escalation plan

Create a clear protocol for handling safety-related incidents that touch customers. For example, if a user reports a serious safety threat to a support agent at 2am, who gets contacted? Define roles and contacts from T&S, CX, Legal, and PR for various scenarios (minor issue vs crisis). Run a drill if possible. Being prepared means faster, calmer responses when real incidents occur.

Share data and define shared KPIs

Swap dashboards on a regular basis. Let T&S see customer satisfaction scores and comments related to safety issues; let CX see volumes and types of content violations or fraud attempts. Define one or two common goals (e.g. "reduce scam-related complaints by 20%" or "improve user-reported content action rate to 95% within 24 hours") and rally both teams to achieve them. Celebrate progress together.

Integrate T&S into product development

Advocate for a seat at the table (or ensure T&S has one) in product design and roadmap discussions. For each new feature or campaign, ask the question: "What could go wrong from a trust perspective, and how will we handle it?" This might mean adding a reporting button, a verification step, or a review process before launch. It's much easier to build it right than to retrofit safety later.

Align on customer communications

Work with Marketing/Communications to review how you talk to users about T&S. Are your help center articles, automated emails, and support scripts conveying the right tone and info? Consistency is key. For instance, if a user gets an account suspension notice, does it provide a path to appeal or contact support? Ensure CX and T&S have input on these templates so they are both firm and user-friendly. Consider proactive comms such as safety tips to users or community updates about policy changes – and coordinate those across teams.

Leverage technology and partners wisely

Explore tools that can ease the load on both CX and T&S. This could be AI moderation software, fraud detection services, or analytics that predict emerging issues. Bring in vendors such as WebPurify or IntouchCX early as part of strategy, not last-minute plug-ins. And when you adopt new tech, involve the people who will work with it daily (moderators, agents) to ensure it truly helps them help customers. A well-integrated tool can improve both safety outcomes and agent experience (which ultimately benefits customers).

Foster a "safety culture" across the company

Encourage cross-team trainings (e.g. a brown-bag session where T&S educates the whole company on how to spot phishing scams, or CX shares top customer pain themes with everyone). Reward collaborative behavior – for example, if a support agent and a policy analyst teamed up to resolve a tough case, recognize that. Make T&S updates a standing item in company all-hands or newsletters. When people see leadership caring about these issues, they'll take initiative too. The goal is to make trust everyone's business.

The platforms that win tomorrow will offer more than just great features; they'll provide experiences that are safer, smoother, and more human. And that starts with making T&S everyone's business.