OASIS CONSORTIUM

Introducing

USER SAFETY STANDARDS FOR OUR DIGITAL FUTURE

USER SAFETY STANDARDS FOR OUR DIGITAL FUTURE

OASIS Consortium is a think tank that includes Trust & Safety experts from community platforms, industry organizations, academia and non-profits, government agencies and advertisers. We're working together to build a better digital world.

We believe that actionable, technology-enabled user safety standards are critical to building ethical brands.

Our user safety standards cover all aspects of building safe, inclusive, and engaged communities.

They follow the Oasis 5Ps of User Safety Framework: Priority, People, Partnership, Product, and Process.

These standards are the foundation of our user safety certification program. Trust & Safety professionals can use them to conduct a self-assessment, validate their work, and secure needed resources.

Metaverse Builders

Industry Orgs.

Academia & Non-Profits

Government

Advertisers

ABOUT THESE STANDARDS

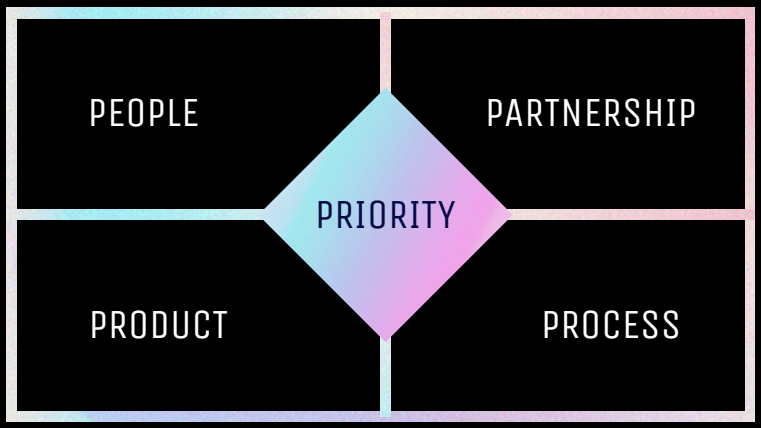
Practical ideas to implement today, based on extensive research

We developed these standards over several months and hundreds of conversations with T&S professionals from across the industry, including gaming, dating and social apps.

- Collected user safety practices from each platform.
- Compared for consistencies & gaps.
- Distilled into Oasis standards.
- Applied across companies to evaluate effectiveness.
- Finalized in collaboration with Oasis advisors.

THE 5P'S OF USER SAFETY FRAMEWORK

All aspects of building safe, inclusive & engaged communities



HOW TO USE THEM

- Share these standards within your company to build support for your T&S initiatives.
- Pledge to uphold the standards and highlight your commitment with the **Oasis User Safety Standards** seal.
- Conduct your internal user safety assessment to measure performance and identify ways to improve.
- Prepare for certification to earn your **Oasis Digital Sustainability in User Safety** seal (coming soon).

USER SAFETY STANDARDS

Start your certification journey to Digital Sustainability in User Safety.

1 PRIORITY

Establish that Trust & Safety is mission-critical for your company.

Accountable Leadership

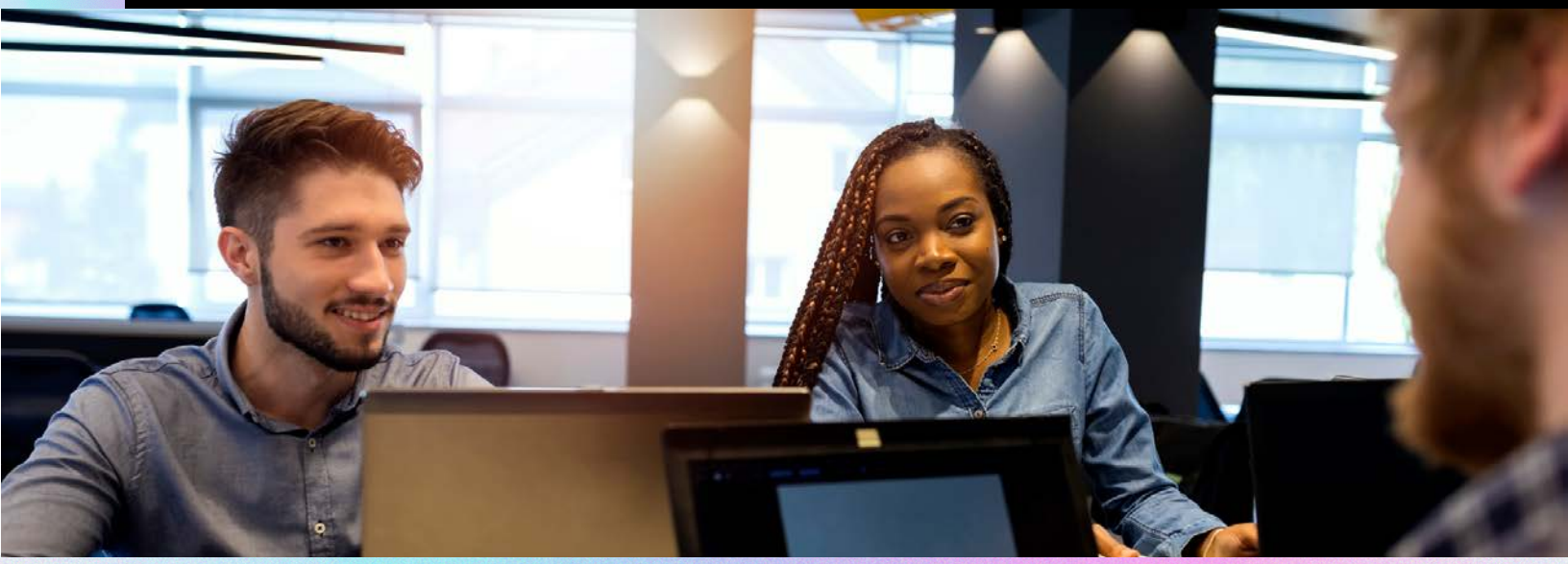
Recognize that user safety is a highest-order, company-wide initiative. Create clarity of ownership by designating an executive-level champion. Establish accountability for both vision and execution. Ensure that T&S leadership is included and supported in product designs and company workflows.

Resources for Development

User safety is a challenge that continues to evolve, along with user behaviors, world events and technical capabilities. Secure resources to invest in a recurring budget. Develop a living roadmap to continue iteration and improvement. Plan for user safety proactively, not reactively.

Cross-Functional Collaboration

Safety-by-Design has implications across Product, Legal, UX and more. Establish top-down management focus on T&S in each discipline. Institute cross-functional working groups. Define T&S goals aligned with the product safety roadmap. Set up mechanisms for follow-through within and across teams over time.



2 PEOPLE

Develop your policies based on representation, learning & wellness.

Diversity & Inclusion

The team that develops policies and enforcement must be diverse in every dimension, especially social background. Select representatives that reflect and understand your community's unique culture. Make sure members are also users.

Learning Organization

T&S professionals track a vast range of information: ever-changing behaviors, regional cultural norms and legal regulations. Become a learning organization to gather best practices from the broader industry. Use findings to inform your innovation initiatives.

Employee Wellness

Moderators and other employees are exposed to the worst of humanity under strict productivity goals. Provide resources and design programs to protect and improve the wellness of your team.

3 PARTNERSHIP

Gain expertise, objectivity & prevention of real-world impacts.

Industry Alliances

Industry associations for T&S professionals are dedicated to best practices and professional growth. Join them to compare notes and grow together. Industry non-profits are staffed by experts in specific human behaviors. Partner with them to access resources to educate users, and research to guide decisions.

Media

Plan regular check-ins with media through your in-house communications teams and external PR agencies. Publicize your commitment and progress toward user safety. This helps foster safety-first brand trust, and build media muscle for crisis management.

Advisory Board

As platforms grow, it makes sense to seek outside opinions on T&S decisions. Institute an independent board of trusted advisors to provide feedback on product updates, policy decisions and gray area cases. This helps to ensure both objectivity and accountability.

Law Enforcement

When platforms identify content that poses a real-world risk, it's essential to notify law enforcement. Partner in advance with all relevant law enforcement agencies, and define appropriate workflows. This may involve a special team called a Law Enforcement Response Team (LERT).

Government Agencies

In addition to local law enforcement, there are legal requirements from international governments and other domestic agencies such as NCMEC. Regulations and workflows vary by group, country and region. Establish contact in advance to understand reporting and record keeping requirements.

4 PRODUCT

Deploy up-to-date technology that enables your success.

Community Guidelines

Set clear expectations for all users, starting with onboarding. A fine-print legal policy with a checkbox is not enough. Feature community guidelines where all users will see them, at key moments and repeatedly over time. Ensure that they are practical to enforce, and regularly updated.

Data for Visibility

Guidelines are useless without enforcement. Data shows what is happening in your community, and whether guidelines are being enforced. Implement tools to collect and analyze data. Define a process to run queries, and develop a data dashboard. Review data regularly to identify trends and inform policy, product and enforcement.

Proactive Detection

Users report only a fraction of violations. Deploy technology that can proactively detect a high proportion of harmful behaviors. Ideally, use contextual AI across all content. Use machine learning models to detect hard-to-identify violations like terrorism, extremism, hate speech and CSAM. Regularly review moderation data and model design.

Moderation Tools

Quality moderation tools can greatly increase moderator effectiveness and efficiency. Select a moderation suite that can automatically create cases, prioritize them, provide context for investigations, assist record keeping and more.

User Reporting Tools

Community guidelines are a commitment to address violations experienced by users. Provide users a mechanism to report violations, and respond promptly. Consider user reports as important feedback for policy design.

5 PROCESS

Define comprehensive processes for an effective operation.

Consistent Enforcement

Community guidelines are the expectations you set for behavior by all users. Enforcement makes guidelines real. Ensure that your enforcement is fairly and consistently applied for all users. This demonstrates your commitment to your values.

Effectiveness Audit

Be certain that your tools are handling the full scope of violations on your platform. Set up an auditing system to regularly measure the performance of your moderation program. Include detection accuracy, automation percentage and response speed.

Hard Cases/ Gray Areas

Borderline cases will always exist. When humans review them, the right answer will not always be clear-cut. Define the process of getting a second opinion or manager approval.

User Appeals

Enforcement decisions are not always correct; every moderation team can make a wrong call. Establish an appeals process to give community members who are punished wrongly an opportunity to overturn the decision.

5 PROCESS, CONT.

Define comprehensive processes for an effective operation.

Escalation	Some violations must be escalated to management or reported to authorities. Define a clear path for each escalation or reporting requirement.
Policy Updates	Review guidelines often, and update any obsolete policies. This helps maintain relevance to evolving online behaviors, real-world events and growing communities.
Bias Prevention	Bias can too easily seep into moderation, negatively impacting already marginalized groups. Actively plan to prevent, detect and remove bias from policies, tools and processes.
Data Security	Platforms must monitor user-generated content to keep users safe. This creates privacy and security risks. Create a framework for T&S, Security and Legal to work together to safeguard data security.
Transparency Reports	Your stakeholders are interested to see how you deliver on your commitment to user safety. Publish transparency reports regularly to share your performance toward T&S goals with users, advertisers, investors and the public.