TaskUs™

Trust & Safety:

Safeguard your platform

Safety that drives Trust & Performance

Digital platforms face unprecedented pressure from surging content volumes, evolving regulatory mandates (like the EU's DSA), and sophisticated bad actors. The Trust & Safety function has become a mission-critical and strategic asset that determines brand trust, user retention, and compliance.

TaskUs delivers an end-to-end Trust & Safety solution that goes beyond basic content moderation. We integrate specialized moderation expertise, digital solutions (including AI safety alignment), and deep psychological care for our teams, helping the world's most innovative companies deliver safety that drives true performance.
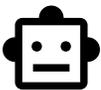
Trust & Safety Operational Challenges

The resilience gap
High burnout and attrition negatively affect moderation teams, leading to inconsistent quality and policy compliance.

The AI risk factor
GenAI introduces new forms of egregious content and model misuse. Without rigorous adversarial testing (Red Teaming), systems risk unpredictable, harmful outputs.

Scaling without sacrificing quality
Rapid user growth, especially in emerging platforms like gaming or social media, requires 24/7 multilingual coverage without inflating operational costs.

Global compliance pressure
Evolving mandates and regional policies require localized expertise, granular reporting, and intelligence-driven threat mitigation, increasing legal and reputational risk.

Evolving regulatory complexity
Constant changes in regional regulations demand immediate policy adaptation, intelligence-driven risk assessments, and robust audit trails, often overwhelming legal teams.

Trust & Safety Solutions

Protecting users & communities
Shielding users and supporting safety across all content formats. We remove harmful content while preserving positive engagement, helping communities thrive.

Securing GenAI systems
Human-in-the-Loop enforcement powers AI models to deliver accurate, ethical outputs. This includes developing Red Teaming prompts to stress-test systems for hidden risks.

Ensuring moderator well-being
Our science-based Wellness-as-a-Service program integrates in-house clinicians and researchers to improve team resiliency, retention, and performance metrics.

Preserving online platform integrity
Reducing risk for marketplaces by verifying identities, moderating listings and ads, and tackling sophisticated fraud and scams that impact revenue streams.

Digital policy & regulatory management
Intelligence-driven insights, regulatory consulting, and risk assessments help platforms stay compliant and mitigate threats before they escalate.

Define the Standard. Control the Risk.

We execute the strategies that leading platforms rely on to maintain their authority and growth. By providing early, actionable intelligence across the ecosystem and removing operational bottlenecks, we ensure your safety program is an accelerator for user adoption and global market entry.

Trust & Safety is AI safety: We establish definitive safeguards for your GenAI models, turning an existential risk into a managed operational function by fusing human oversight with machine learning.

Wellness drives performance: Our clinical model treats moderator resilience as a non-negotiable performance metric, ensuring high quality through continuous, trauma-informed support.

Platform innovation requires scale: We provide the global, proprietary tools and 24/7 multilingual talent needed to meet explosive growth without operational failure.

Proactive regulatory compliance: We provide the intelligence and audit trails to ensure compliance before the mandate hits.

Why Partner with Us?

Trusted & tested partner

Specialized talent

Human-in-the-Loop expertise

Proven results

TaskUs™

🌐 www.taskus.com

📞 (888) 400-TASK