Content Moderation Technology Trailblazers: Top Technology Start-ups Creating Buzz in the Market

November 2024: Complimentary Abstract / Table of Contents

Market Report

Trust and Safety

Everest Group®

Our research offerings

This report is included in the following research program(s):

Trust and Safety

- Advanced SciTech
- Amazon Web Services (AWS)
- Application Services
- Artificial Intelligence (AI)
- Asset and Wealth Management
- Banking and Financial Services Business Process
- Banking and Financial Services Information Technology
- Catalyst™
- Clinical Development Technology
- Cloud and Infrastructure
- Contingent Staffing
- Contingent Workforce Management
- Contract Research Organization Services
- Customer Experience Management Services
- CX Excellence
- CXM Technology
- Cybersecurity
- Cyber Threat Detection and Response
- Data and Analytics
- Digital Adoption Platforms
- Digital Services
- Digital Workplace
- Employee Experience Management (EXM) Platforms
- Employer of Record (EOR)
- Engineering Research and Development
- Enterprise Platform Services
- Exponential Technologies

- Finance and Accounting
- Financial Crime and Compliance
- Financial Services Technology (FinTech)
- Forces & Foresight
- GBS Talent Excellence
- Global Business Services
- Google Cloud
- HealthTech
- Human Resources
- Insurance Business Process
- Insurance Information Technology
- Insurance Technology (InsurTech)
- Insurance Third-Party Administration (TPA) Services
- Intelligent Document Processing
- Interactive Experience (IX) Services
- IT Services Excellence
- IT Talent Excellence
- Lending and Mortgages
- Life Sciences Business Process
- Life Sciences Commercial Technologies
- Life Sciences Information Technology
- Locations Insider™
- Marketing Services
- Market Vista™
- Microsoft Azure
- Microsoft Business Application Services
- Modern Application Development (MAD)

- Multi-country Payroll
- Network Services and 5G
- Oracle Services
- Outsourcing Excellence
- Payer and Provider Business Process
- Payer and Provider Information Technology
- Payment Integrity Solutions
- Price Genius – AMS Solution and Pricing Tool
- Pricing Analytics as a Service
- Process Intelligence
- Process Orchestration
- Procurement and Supply Chain
- Recruitment
- Retail and CPG
- Retirement Technologies
- Revenue Cycle Management
- Rewards and Recognition
- SAP Services
- Service Optimization Technologies
- Software Product Engineering Services
- Supply Chain Management (SCM) Services
- Sustainability Technology and Services
- Talent Genius™
- Technology Skills and Talent
- Trust and Safety
- Value and Quality Assurance (VQA)

If you want to learn whether your organization has a membership agreement or request information on pricing and membership options, please contact us at info@everestgrp.com

Learn more about
our custom research capabilities

Benchmarking

Contract assessment

Peer analysis

Market intelligence

Tracking: providers, locations, risk, technologies

Locations: costs, skills, sustainability, portfolios

Contents

For more information on this and other research published by Everest Group, please contact us:

David Rickard, Partner

Abhijnan Dasgupta, Practice Director

Dhruv Khosla, Senior Analyst

Tushar Pathela, Senior Analyst

Neha Sharma, Manager

Muskan Negi, Research Associate

Background of the research

The exponential growth of user-generated content, the rise of misinformation, and the constant threat of harmful content present formidable challenges for platforms that human moderation alone cannot address. An effective technology-based solution can fast-track the moderation process and minimize the requirement for a moderator going through the toxic content. As online platforms continue to grapple with harmful content, they are increasingly investing in automated tools to prevent toxic user experiences.

The current content moderation technology landscape consists of enterprises developing their own in-house technology and third-party providers supplying technology solutions to these organizations. The third-party technology landscape allows enterprises to leverage advanced tools and specialized expertise for managing complex moderation needs and adapt to evolving regulatory requirements, all without the need to build and maintain in-house solutions. However, wider adoption of third-party content moderation technology solutions needs to overcome data privacy concerns associated with sharing user data and extended timeframes required to achieve Return on Investment (RoI).

The current third-party technology landscape comprises established legacy companies (holding approximately 67% of the market share) and **emerging start-ups (focus of the report)** providing diverse solutions across ML classifiers, APIs, and platforms to help enterprises detect, moderate, and report harmful content. Beyond developing these solutions, these vendors often collaborate with each other and regulatory bodies to foster co-innovation, enhance solution effectiveness, and streamline moderation and reporting processes.

In this research, we assess 13 leading content moderation technology provider start-ups, primarily focusing on their core capabilities and market adoption and domain coverage. We present an assessment, detailed profiles, and Everest Group's Point of View (POV) of trailblazers in the content moderation technology market.

The assessment is based on primary interactions with select organizations coupled with Everest Group's ongoing tracking of the content moderation technology landscape.

Scope of this report

Geography: Global

Industry: All

Services: T&S

Use cases: leading start-ups in the content moderation technology market

Based on the assessment framework, we identified 13 of the shortlisted 27 high-potential start-ups as trailblazers, after an in-depth dimensional analysis
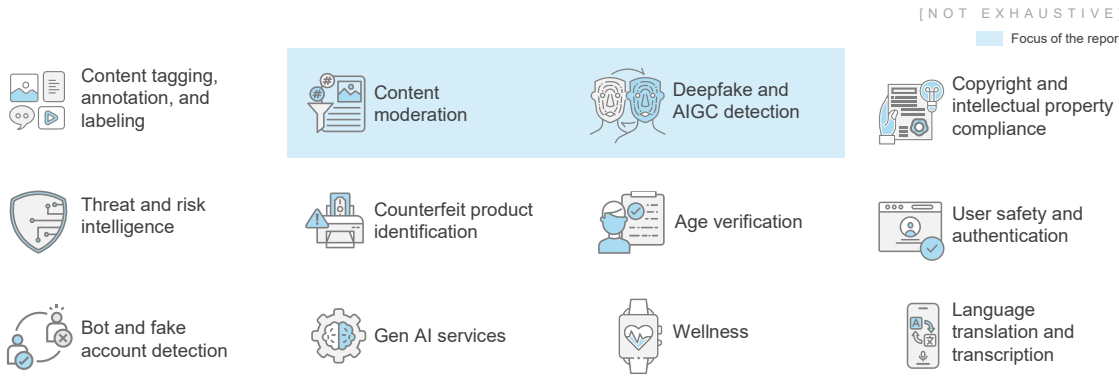

13 identified as trailblazers

This study provides a deep dive into leading content moderation technology provider start-ups; below are four charts to illustrate the depth of the report
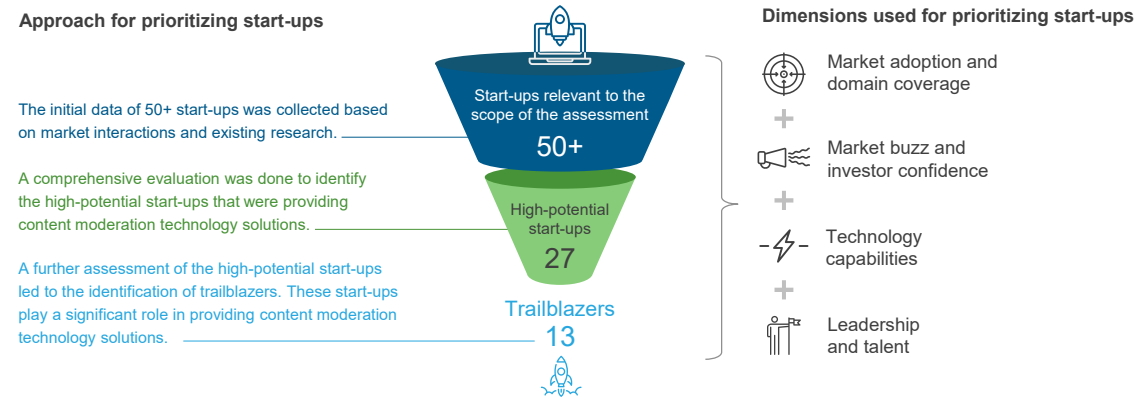
Different technology solutions within T&S

[N O T   E X H A U S T I V E]

▢ Focus of the report

- Content tagging, annotation, and labeling
- **Content moderation**
- **Deepfake and AIGC detection**
- Copyright and intellectual property compliance
- Threat and risk intelligence
- Counterfeit product identification
- Age verification
- User safety and authentication
- Bot and fake account detection
- Gen AI services
- Wellness
- Language translation and transcription

Key offerings

Providers with API-focused offering

Core focus
Offer modular solutions with pre-defined sub-abuse policies that enterprises can integrate into their existing systems, allowing them to assign risk scores to content based on established guidelines

Support offered
AI models that detect threat at abuse (such as hate speech, violence, and adult content and nudity) and sub-abuse levels

Value proposition:
- Offer comprehensive support for detection of a particular threat type(s)
- Easy integration with existing infrastructure and workflows

Suitability:
Enterprises having existing tech stack and looking for specialized models for particular threat types

Providers with platform-focused offering

Core focus
Offer end-to-end content moderation solutions that bundle various features such as moderation, reporting, and user management into a unified platform

Support offered
Platform with support for moderation (via own APIs or third-party integrations), content queuing, analytics on policy violations and moderator performance, community flagging, and tools for wellness

Value proposition:
Centralized management for all content moderation workflows to reduce complexity

Suitability:
Enterprises looking for all-in-one solution with minimal development overhead and centralized management

Examples of providers

HIVE   MODULATE   THORN   TRUSTLAB   ActiveFence   checkstep   CINDER   TREMAU

Assessment methodology

Approach for prioritizing start-ups

The initial data of 50+ start-ups was collected based on market interactions and existing research.

A comprehensive evaluation was done to identify the high-potential start-ups that were providing content moderation technology solutions.

A further assessment of the high-potential start-ups led to the identification of trailblazers. These start-ups play a significant role in providing content moderation technology solutions.

Start-ups relevant to the scope of the assessment
50+

High-potential start-ups
27

Trailblazers
13

Dimensions used for prioritizing start-ups

- Market adoption and domain coverage
+
- Market buzz and investor confidence
+
- Technology capabilities
+
- Leadership and talent

Source: Everest Group (2024)

Trailblazers in content moderation technology profiles

Company overview
- Provider 1 provides content moderation solutions via AI and human experts to ensure cultural sensitivity and brand alignment
- Its solutions supports moderation across multiple subclasses including sexual content and nudity, harassment and hate speech, spam and phishing, violence, and child safety

Headquarters: New York City, New York   **Founding year:** 2021

Fact sheet

Key leaders	• ABC, Co-founder and Chief Executive Officer • XYZ, Co-founder and Chief Technology Officer
Illustrative clientele	Client 1, Client 2, Client 3, and Client 4
Partnerships	Partner 1, Partner 2, Partner 3, and Partner 4
Recent announcements	• August 2024: XYZ • March 2024: XYZ
Key investors and funding	In 2021, raised US$XX million in funding from Investor 1, Investor 2, and Investor 3

Market adoption and domain coverage	Market buzz and investor confidence	Technology capabilities	Leadership and talent	Overall

Content moderation offering

Key offerings include solution 1, solution 2, and solution 3
- Solution 1 leverages deep learning, image and voice recognition, and ML models to identify and moderate content across abuse areas and content types
- Solution 2 allows to configure tools to moderate content and facilitates user reports and auditing suspicious actors
- Solution 3 conducts user behavior analysis with risk assessments and enables real-time identification of CSAM and terrorist content

Everest Group's POV
- **Differentiators:** Player 1 offers support across all content types and caters to clients in the social media, dating, streaming, and gaming industries
- **Areas of improvement:** its solution does not support detection and moderation of deepfakes and AIGC

Everest Group®

Research calendar
Trust and Safety

Published   Current release   Planned

Reports title	Release date
R'AI'sing the Bar: The Enhanced Role of AI in Ensuring Safe Digital Spaces in the Future	January 2024
The Morphing Face of Threats	February 2024
Trust and Safety Services PEAK Matrix® Assessment 2024	March 2024
Preserving the Ballot: The Role of Trust and Safety in Election Misinformation	May 2024
Trust and Safety Services – Provider Compendium 2024	May 2024
Ordinance at the Gates: the Evolving Landscape of AI Regulations Around the World	May 2024
Gen AI Impact and Adoption in Trust and Safety (T&S)	June 2024
User Safety Redefined in the Generative AI Era: Trust and Safety State of the Market 2024	August 2024
Under the Lens: The "Trust" of Transparency Reports	November 2024
Content Moderation Technology Trailblazers: Top Technology Start-ups Creating Buzz in the Market	November 2024
Key Tech Away 2.0	Q4 2024
Trust and Safety Chronicles: Unpacking 2024, Anticipating 2025	Q1 2025
Age Gate Unlocked: Navigating Challenges and Innovations in Age Verification	Q1 2025
Trust and Safety Services PEAK Matrix® Assessment 2025	Q1 2025
Beyond the Cart: Safeguarding the E-Commerce Ecosystem	Q2 2025
Trust and Safety Services – Provider Compendium 2025	Q2 2025

Note: Click to see a list of all of our published Trust and Safety reports

Stay connected

Dallas (Headquarters)
info@everestgrp.com
+1-214-451-3000

Bangalore
india@everestgrp.com
+91-80-61463500

Delhi
india@everestgrp.com
+91-124-496-1000

London
unitedkingdom@everestgrp.com
+44-207-129-1318

Toronto
canada@everestgrp.com
+1-214-451-3000

Website
everestgrp.com

Blog
everestgrp.com/blog

Follow us on

in X ▶ ⊙ f

Everest Group®
With you on the journey