NISOS

Trust & Safety Buyer's Guide

Protecting Online Platforms with Threat Intelligence

Introduction

As a member of the Trust and Safety team, you play a critical role in creating and enforcing policies that govern your organization's products and services. Success has never been more critical as online platforms are intrinsic to how we communicate, develop communities, and grow our businesses worldwide.

However, these platforms increasingly serve as a launching pad for malicious activity and a hunting ground for cyber predators, bringing unintended consequences and risking your brand.

Today, Trust and Safety teams don't just identify and remove harmful content; they are intimately involved in developing tools, systems, and techniques that ensure compliance with platform policies. As such, you sit at the crossroads, skillfully balancing the objectives of protecting users, supporting platform partners, empowering creators, and courting advertisers, while ensuring the sustainable growth of the platform.

The organization looks to you to keep the lights on in the face of increasingly sophisticated attacks, fraud, the spread of misinformation and disinformation, and rapidly shifting global regulations around online content and user privacy.

This guide will explore how Trust and Safety teams can improve protection for their platform, brand, and users with Open Source Threat Intelligence.

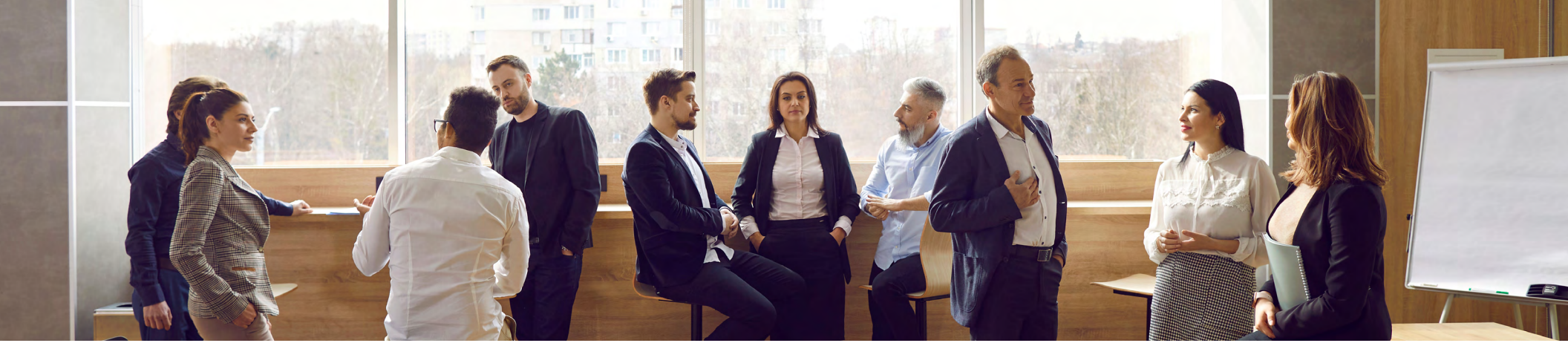
Threats to Online Platforms

Trust and Safety teams are charged with creating an environment that enables users to have the best experience possible with a product or service. Establishing and enforcing acceptable use policies and preventing, detecting, and responding to the abuse that occurs on your platform is a growing challenge. Failing to act threatens to erode user experience, damage your brand, and take a bite out of profits.

Ensuring online communities are free of abuse, criminal activity, and fraud requires persistence. Ill-intentioned users, fraudsters, cybercriminals, and even nation-state actors are constantly finding new ways to exploit vulnerabilities in online platforms.

Threats are more prevalent than ever, varying from disinformation, platform abuse, brand dilution, strategic breach campaigns, extortion, insider threats, and intellectual property theft. Threats against online platforms generally fall into one of three categories: content violations and misuse, fraud, or platform abuse.

$25 Billion	$2,436	36%	35%
total cost of account takeover attacks in 2020 [1]	average value of a fraudulent transaction in 2021 [2]	of data breaches caused by phishing attacks [3]	of all platform traffic is from bad bots [4]



Top Responsibilities of Trust and Safety Teams

Policy and Enforcement	Establishing policies that prohibit stalking, doxxing, or other forms of harassment or violence and taking action against users who violate them, such as suspending or banning their account.
Reporting and Support	Providing users with the ability to report incidents, offering support and resources to affected users including information about how to contact law enforcement or other relevant authorities.
Technology and Tools	Deploying the technology and tools needed to detect and prevent harmful behavior, and overseeing moderation tools to remove harmful content or block users who engage in harmful behavior.
Education and Awareness	Educating users about the risks of stalking, doxxing, or violence, and providing resources and guidance on how to stay safe online, including how to recognize warning signs of potential scams.
Collaboration with Law Enforcement	In cases where physical harm is threatened or has occurred, Trust and Safety teams collaborating with law enforcement to investigate and pursue legal action against perpetrators.

Challenges to Protecting the Integrity and Trust of Online Platforms

Identifying problematic content concerning behavior, threats, and illegal activity on your platform is a time-intensive process. Sifting through user reports and alerts generated by automated tools makes it difficult for even the most advanced team to stay on top of misuse and criminal behavior on their platforms. As a result, 51% of Trust and Safety leaders feel their teams are reaching burnout, citing the speed of threats and a lack of executive support to be driving stressors [5]. As your platform grows, the challenge becomes even greater.

20%	50%	$3.56 Billion	80%
Of consumers who experienced a false positive abandoned the platform.[6]	The share of Americans who get misinformation on social media. [7]	The amount American consumers reported lost due to online fraud during 1H 2022. [8]	The percent of people who encounter a scam on social media. [9]

5 Considerations for Building a Threat Intel Program to Support Trust and Satefy Teams

- **Keeping up with the constantly evolving threat landscape:** Threats emerge and evolve rapidly, making it difficult for even mature teams to keep up. Continuously monitoring and analyzing various sources of information to identify emerging threats and stay ahead of attackers requires a significant investment in technology, tools, and skilled personnel.
- **Team and constraints:** With economic woes on the horizon, Trust and Safety programs must do more with less. Hiring for Trust and Safety teams has seen cost-cutting across the board, and hiring for roles is down 70% year over year.
- **Limited resources:** Building and maintaining a threat intelligence program requires significant time, money, and resources. Smaller organizations or those with limited budgets may struggle to build an effective threat intelligence program.
- **Privacy concerns:** Collecting and analyzing data to identify threats may run afoul of user privacy concerns. Organizations must balance the need for privacy and security while maintaining trust with their users.
- **Integration with existing workflows:** Threat intelligence programs must integrate with existing workflows to ensure that relevant information is shared with the appropriate teams, requiring close collaboration with incident response, legal, and communications, and other stakeholders.

Open Source Threat Intel for Platform Protection

Open Source Intelligence (OSINT) refers to collecting, analyzing, and disseminating information from publicly available sources such as social media, news articles, government reports, and websites. For Trust and Safety teams, OSINT can provide valuable insights into online activities that may pose risks to users or the platform.

By monitoring public sources of information, such as social media, forums, and other online communities, OSINT can help identify patterns of behavior that may indicate fraudulent activity, hate speech, harassment, or other violations of platform policies.

The Importance of the Dark Web

The dark web is a part of the internet not indexed by search engines and requires specific software, such as the Tor browser, to access. This anonymity makes it a breeding ground for illegal activities. Maintaining visibility in the dark web makes it possible to identify platform breaches, track the origin of platform abuse, anticipate cyber-attacks, and stay ahead of shifting sentiments around your brand in the cybercriminal underground.

Use Case: Content Moderation

As your platform grows and the volume of content and users increases, content moderation becomes increasingly challenging. Content violations can lead to user dissatisfaction, harm the platform's reputation, and create legal liability for the platform. Examples of content violations include hate speech, harassment, pornography, and illegal activities. Monitoring and investigating many user reports and alerts about potentially problematic content, behavior, or activity on the platform is a growing issue. Even with the host of tools available to help you automate content classification and analysis, acting on this information alone can risk the trust of your users and damage your brand.

The onus is on your team to collect as much information as possible about the concerning content, including the type of content, its context on the platform, and any relevant policy violations. While the data derived from your platform and supporting infrastructure can help, bringing context from outside the platform and organization ensures you have a complete understanding of any issue.

Task	Intelligence Support
Identifying fake and fraudulent accounts	Patterns and similarities in account creation, including using similar email addresses or IP addresses, accounts created on the same day, etc., can be signs of a coordinated effort to create fake accounts.
Investigating the likelihood of a threat	Assessing the likelihood of threats of violence or other types of harm by investigating the offending user's history, including presence on social media, helps to determine the credibility and capability of a threat actor.
Attributing sources of harmful content	Tracking the origin of harmful content to its source allows you to take appropriate action against the users who created the content, preventing it from spreading further.
Verifying user-generated content	Using reverse image search tools or analyzing metadata can help determine user-generated content's source, authenticity, and context.
Evaluating the reputation of third-party sources	Keeping users safe from disinformation by evaluating news outlets, websites, or blogs sharing harmful content by analyzing the content, authors, and sources of information.

Use Case: Fraud

Fraudsters are constantly developing new tactics and methods to evade detection, making it difficult for Trust and Safety teams to keep up. Fraud involves intentional deception or misrepresentation to exploit users for personal or financial gain. Examples of fraud on online platforms include phishing scams, identity theft, fake profiles or reviews, and fraudulent transactions. Fraud can erode user trust, harm users financially, and damage the platform's reputation.

Missing signs of fraud can lead to unintended consequences in attrition and can have a devastating impact on your users and brand. At the same time, strong fraud prevention measures can sometimes create friction for users, such as when they must provide additional verification or authentication steps. Overly aggressive fraud prevention measures can also lead to false positives, where legitimate users are mistakenly flagged as fraudulent.

Task	Intelligence Support
Investigating fraudulent identities	Verifying the identities of users on the platform and cross-referencing other personal information to ensure it matches what's publicly available online to prevent fraudsters from using fake identities.
Tracking fraudulent transactions	Searching the dark web for relevant mentions of stolen credit card numbers, credentials, and other sensitive information that can be used in fraudulent transactions.
Investigating phishing attacks	Investigating phishing attacks by analyzing the URLs and email addresses and cross-referencing this information with publicly available data and breach datasets.
Monitoring for scams	Monitoring social media, forums, and other online platforms for suspicious activity and keywords related to scams targeting platforms like yours.
Monitoring for brand abuse	Monitoring social media and other online platforms for mentions of the company's name or logo to identify brand impersonation and abuse.

Use Case: Platform Abuse

Cybercriminals often leverage legitimate platforms to facilitate and organize their attacks. Platform abuse refers to any type of behavior that seeks to exploit the platform's systems or services for malicious purposes. Examples of platform abuse include spamming, hacking, distributing malware, and creating bots to manipulate platform activity. Platform abuse can disrupt platform operations, harm user experience, and create security risks for users.

Using commercially available platforms lowers the barrier to entry for phishing and other attacks, providing ready infrastructure to host fraudulent pages and content. Worse still, exploiting a legitimate platform makes it possible for malicious content to appear legitimate, defeating consumer cyber protections.

Task	Intelligence Support
Social media monitoring	Looking for mentions of your platform or brand that could indicate phishing attempts or bot activity, especially those that include suspicious URLs or links shared en mass.
Domain analysis	Check the domain registration information and see if it matches the organization's information and whether a domain associated with a suspicious link or email is legitimate.
Image analysis	Performing reverse image searches to identify where profile pictures and other images are used across multiple accounts or taken from a stock photo website.
Malware analysis	Performing malware analysis to determine the source of a suspicious link or attachment and whether it contains any malware or virus.
Dark web monitoring	Searching the dark web for stolen credentials, user data, bots, or exploits offered for sale that purport to target your platform or brand.

Defining Stakeholder Intelligence Needs

Intelligence programs thrive when they are aligned with the actions a stakeholder could take based on the intelligence they receive. The typical Trust and Safety organization includes distinct teams for policy and enforcement, each with different objectives but supporting the overall goal of platform protection.

Trust and Safety teams investigate issues on their platform by following a set of procedures designed to identify, evaluate, and address potentially harmful content, behaviors, or activities. Throughout the process, Trust and Safety teams may collaborate with other teams, such as legal, engineering, or customer support, to ensure that the appropriate actions are taken and that users are kept safe.

Intelligence stakeholders in the context of Trust and Safety generally fall into the following categories:

Policy Teams	Drive the strategy, vision, and execution for preventing and reducing policy-violating content and behavior on the platform.
Enforcement Teams	Responsible for day-to-day services, including moderation activities, incident management, and responding to issues affecting a particular product or service.
Developers	Engineering team that develops the tools used by content moderators and implements new product features to mitigate harmful experiences, educate users about policies, and improve users' trust.
Knowledge Management	Support content moderation teams with training resources on policy application and champion user policy education.
Compliance and Legal Teams	Responsible for reviewing and accurately assessing legal requests from law enforcement, ensuring compliance, and working closely with internal public policy and legal teams.

Establishing Your Threat Intelligence Workflow

Assess, Monitor, and Investigate

Threat Landscape Assessment | Understand your organization's complete risk profile - across multiple security domains - and receive specific platform risk reduction guidance

The foundation of a successful Trust and Safety program starts with a comprehensive, repeatable evaluation of your organization's threat landscape. A threat landscape assessment provides a clearer picture of your platform and organization's key threats, vulnerabilities, and exposure. This will provide an intimate understanding of your organization by pinpointing critical assets and connecting them to specific threats and scenarios. This way, you can align your resources with the right risks and avoid wasting time and resources on low-priority threats.

Threat / Issue Monitoring | Stay on top of your organization's specific risks before they become threats

With a threat landscape assessment serving as a baseline, monitoring the surface, deep, and dark web for mentions about a company, brand, product or service, and keeping tabs on target users and trends ensures you remain proactive. Establishing a monitoring capacity doesn't stop with tool selection. Developing and refining feeds, integrating them into your systems, and establishing the scope of monitoring takes time.

Investigations and Requests for Information | Bolster your team's numbers and capabilities

Mature organizations understand that the more they learn, the more questions they have that require further investigation and analysis. Driven by Requests for Information (RFIs), investigations are arguably the most crucial process in the threat intelligence lifecycle as they allow a deeper look into specific threats or concerns and questions from key stakeholders.

Attribution and Unmasking: When Is It Appropriate?

Attribution and unmasking are the strongest mechanisms for deterrence to get malicious activity to cease. If your organization faces an ongoing, unresolved, or recurring threat to your platform or users, attributing or unmasking the threat actors behind the scheme can provide immediate value.

With a clear sense of who is behind an issue, including how and why attackers are targeting your organization, it is important to implement the proper technical controls.

In cases of economic or business impact, stronger measures of attribution and unmasking may be necessary to pursue civil lawsuits or retribution against named individuals.

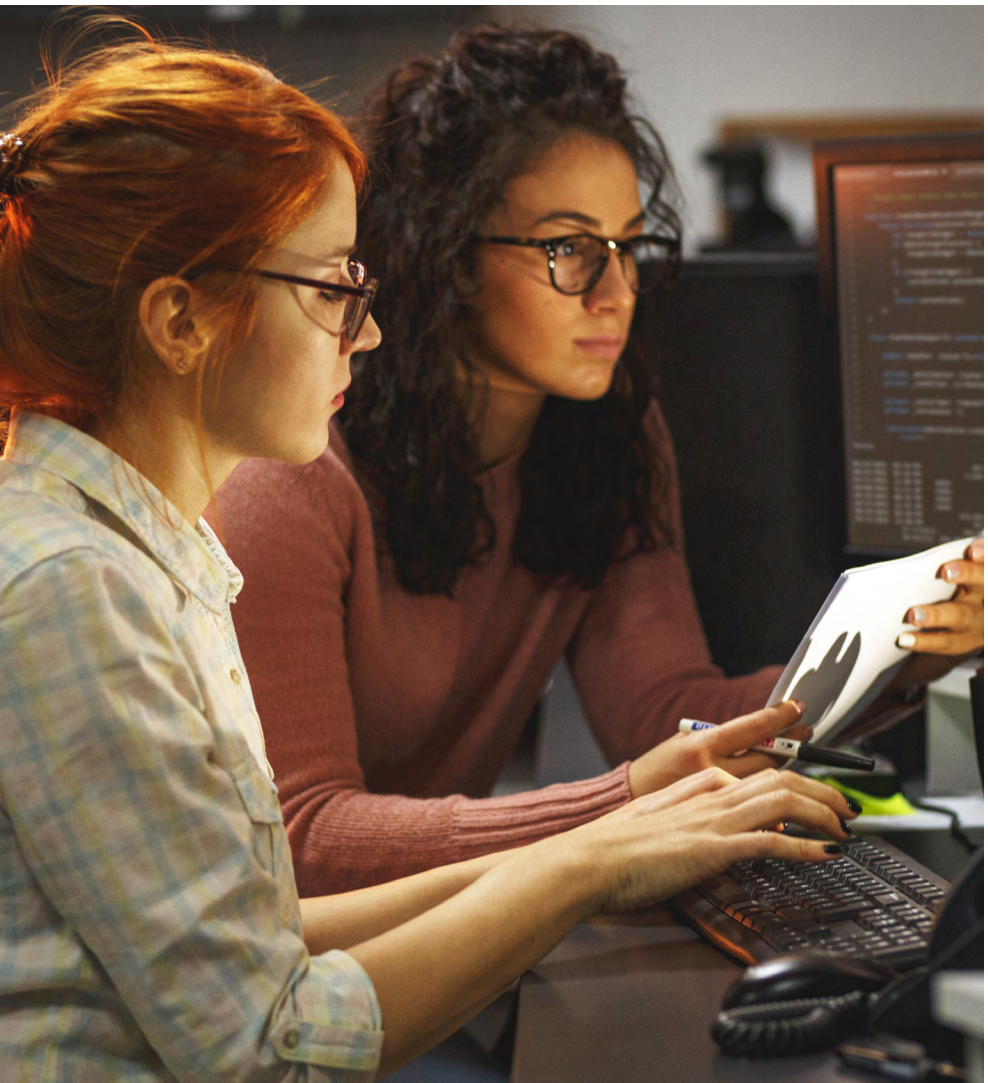
Attributing the threat actor group responsible gives you better insight into the actor's intentions with the stolen data. Attribution can help create a defensible narrative and clearly justify security measures to regulators, users, and other stakeholders.

Some Examples of the Working Effectively are:

- Contacting the perpetrator's family members or employer to intervene
- Law enforcement conducting a "knock and talk" without prioritizing prosecution
- Rolling back anonymity by filing civil lawsuits and sending cease and desist letters
- Working with law enforcement to prioritize prosecution

Intelligence: A Natural Fit for Managed Services

Developing threat data into actionable intelligence takes time, skill, experience, as well as the right tools. Enterprise security teams spend the majority of their cycles dealing with raw data and reviewing pre-populated threat dashboards, which keep them from effectively and proactively investigating risks to their organization. Investigations, as a result, end up being shallow, as intel teams simply lack the time to properly evaluate and analyze each critical alert they receive.



While these threat data feeds and platforms provide value, they fail to meet the business's unique needs and deliver intelligence. Many intelligence products or feeds available in the market provide unfinished intelligence, only providing organizations with a generalized piece of the picture and failing to deliver business-specific actionable outcomes.

Managed Intelligence providers fill a crucial gap by combining people, processes, and technology to deliver threat intelligence as a service.

What to Look for in a Managed Intelligence™ Provider

Threat intelligence is a critical element of any serious security strategy, but few security teams have the expertise or resources to tackle all the threats they face. Managed Intelligence providers fill a crucial gap by combining people, processes, and technology to deliver threat intelligence as a service.

A Managed Intelligence Provider allows organizations to offload resource-intensive threat intelligence tasks to an experienced partner provider.

7 Things Managed Intel Providers Should Do

1. Generate intelligence specific to your organization
2. Deliver analyst-led finished intelligence with access to the analysts
3. Utilize multi-source collection and analysis capabilities
4. Leverage multilingual data sources and analysis
5. Discover and understand the adversarial mindset (motivations and intended outcomes)
6. Attribute and unmask adversaries based on relevance and need
7. Provide intel advice and threat actor engagement guidance

Nisos: The Managed Intelligence Company™

For enterprise security teams with tight budgets, limited time, and expertise in short supply, Nisos fills a crucial gap by combining people, processes, and technology to deliver threat intelligence as a managed service. Nisos experts monitor, identify, analyze, and investigate risks to provide client-specific intelligence that is necessary to stop threats.

Unlimited Access, Unlimited Questions

The Nisos Managed Intelligence™ Suite allows you to offload complex threat intelligence efforts to an expert analyst team focused on your needs. Nisos analysts have the tools and experience to efficiently reveal critical open-source intelligence from the surface, deep, and dark web to identify threats in your security shadows.

Threat Landscape Assessment	Managed OSINT Monitoring	Adversary Insights® Investigations	Executive Shield Digital
Comprehensive baseline assessment of your organization's threat profile	External threat monitoring, investigation, and critical threat alerting	Analyst expertise to identify and investigate risks and counter adversary threats	Digital risk assessment, monitoring, plus PII identification and removal

A Partner Focused on Your Intelligence Needs

Working as an extension of your team, Nisos provides intelligence focused on real-world threats specific to your organization. With Nisos as a partner, you can be confident in your ability to respond to advanced threats, even as your team evolves. You benefit from our broad experience and extensive toolset, so you'll always have the resources to fill knowledge gaps and address unique stakeholders' needs. Nisos analysts work with your team to respond to Requests for Information (RFIs) on your most pressing security concerns and support ongoing security operations with monitoring and alerts.

Reasons to Partner with Nisos for Threat Intelligence

1. Unmatched Collection Capabilities

Using an integrated toolset of over 30 third-party and proprietary tools, Nisos collects and maintains a vast collection of content to query evidence of exposure and keep tabs on your threats.

2. Team Pandion®

Pandion, our team of elite intelligence analysts, average 10+ years of US Intelligence Ops and Fortune 500 experience. They provide unmatched cross-functional expertise and insights into adversarial challenges.

3. Extension of Your Team

With Nisos, you work with named technical operators and analysts who contextualize their findings. Engagement is scoped to your needs.

4. Closed Forums

Using appropriate tradecraft and following legal guidance, Nisos is uniquely able to access closed cybercriminal forums, and connect with persons of interest, including threat actors, to obtain insights important to you.

5. No Noise

Nisos doesn't provide a feed or stream of alerts you'll have to silence. We only alert you to issues you should address.

6. Analyst Engagement and Client Success

Nisos experts are at the center of each engagement. As a Nisos client, you have access to a Lead Analyst and a Client Success Director who are focused on your ongoing intelligence needs.

7. Right-Sized Reporting

Detailed reports with recommendations that include prioritized actions, next steps, and key considerations specific to each client.

8. Immediately Useful Intelligence

Nisos delivers finished intelligence, not just a statement of facts. A report from Nisos provides real answers, to quickly understand the who, what, when, and how behind everything we uncover.

Sources:

1. https://cybersecurityventures.com/cybercrime-damage-costs-6-trillion-by-2021/

2. https://www.sift.com/blog/the-2021-digital-trust-insight-report

3. https://enterprise.verizon.com/resources/reports/dbir/

4. https://www.imperva.com/blog/the-2021-bad-bot-report/

5. Vanson Bourne 2022

6. https://blog.sift.com/customer-insult-survey/

7. Pew Research Center survey Jul-Aug 2022

8. Atlas VPN, August 2022

9. Social Media Today, 12/5/22

Explore Nisos

Analyst-Led Threat Intelligence

Nisos is The Managed Intelligence Company™.

Our services enable security, intelligence, and trust & safety teams to leverage a world-class intelligence capability tailored to their needs.

We fuse robust data collection with a deep understanding of the adversarial mindset delivering smarter defense and more effective response against advanced cyberattacks, disinformation and abuse of digital platforms.

For more information visit www.nisos.com or email info@nisos.com